Thesis/
Reports
Ajzen,
I.

INCENTIVE COMPATIBILITY IN ASSESSING THE VALUE OF PUBLIC GOODS

Incentive Compatibility in Assessing the Value of Public Goods

Icek Ajzen
Department of Psychology
University of Massachusetts
Amherst, Massachusetts

and

Thomas C. Brown
Rocky Mountain Research Station
USDA Forest Service
Fort Collins, Colorado

Final Report
RMRS-99190-RJVA

between
University of Massachusetts
and
Rocky Mountain Research Station

December 2004

ROCKY MOUNTAIN RESEARCH STATION LIBRARY
FORT COLLINS CENTER
240 W PROSPECT RD
FORT COLLINS, CO 80526-2098

This final report consists of the following papers:

- Brown, T. C., I. Ajzen, and D. Hrubes. 2002. Obtaining unbiased contingent values: Further tests of entreaties to avoid hypothetical bias. DP-01-1, RM-4851, Rocky Mountain Research Station, Fort Collins, CO.
- Brown, T. C., I. Ajzen, and D. Hrubes. 2003. Further tests of entreaties to avoid hypothetical bias in referendum contingent valuation. *Journal of Environmental Economics and Management* 46(2): 353-361.
- Ajzen, I., T. C. Brown, and F. Carvajal. 2004. Explaining the discrepancy between intentions and actions: The case of hypothetical bias in contingent valuation. *Personality and Social Psychology Bulletin* 30(9): 1108-1121.

Discussion Paper DP-01-1 revised, RM-4851

Obtaining Unbiased Contingent Values: Further Tests of Entreaties to Avoid Hypothetical Bias

Thomas C. Brown, Icek Ajzen, and Daniel Hrubes[1]

1 March 2002

Abstract

Over-estimation of willingness to pay in contingent markets has been attributed largely to hypothetical bias. One promising approach for avoiding hypothetical bias is to tell respondents enough about such bias that they self-correct for it. A script designed for this purpose by Cummings and Taylor (1999) was used in hypothetical referenda that differed in payment amount and in secrecy of the vote. In comparisons with behavior observed in otherwise identical real payment referenda, the script worked well whether the vote was open or secret at higher payment levels, but inadequately corrected the bias at a lower payment level. Generally, both hypothetical bias and the effectiveness of the script in correcting for that bias increased as did referendum payment amount.

Introduction

Given the goal of efficient management of environmental resources but the difficulty of estimating the benefits of environmental programs using revealed preference methods, hopes have been high that properly designed contingent markets would emulate actual markets. However, evidence has mounted that willingness to pay measured in contingent markets over-estimates payment observed in otherwise identical real cash markets. Differences between real and hypothetical willingness to pay, which we will call hypothetical bias, have been found for private goods such as a chocolate bar (Kealy, Dovidio, & Rockel, 1988), a houseplant (Boyce, Brown, McClelland, Peterson, & Schulze, 1989), an original landscape painting (Neill, Cummings, Ganderton, Harrison, & McGuckin, 1994), a juicer (Cummings, Harrison, & Rutstrom, 1995), an art print (Loomis, Brown, Lucero, & Peterson, 1996), and a sports card (List, 2001). Significant over-

[1] Brown is with the Rocky Mountain Research Station, U.S. Forest Service, Fort Collins, Colorado. Ajzen is with the Department of Psychology, University of Massachusetts, Amherst. Hrubes was at UMass but is not with the Department of Psychology, College of Mount Saint Vincent, Riverdale, New York.

estimations have also been found for funding public goods such as removing roads to enhance wilderness character along the North Rim of the Grand Canyon (Brown, Champ, Bishop, & McCollum, 1996), distributing a citizens' guide to households affected by groundwater contamination in Albuquerque (Cummings, Elliott, Harrison, & Murphy, 1997), and helping the Nature Conservancy protect rain forest in Costa Rica (Cummings & Taylor, 1999). These studies used a mixture of bidding mechanisms, suggesting that the over-estimation cannot be ignored.[2]

Approaches for accurately estimating willingness to pay in contingent markets include (1) using real payment bids for other goods as a way to calibrate the hypothetical bids (Blackburn, Harrison, & Rutstrom, 1994; Fox, Shogren, Hayes, & Kliebenstein, 1998), (2) asking respondents how certain they are that they really would be willing to pay as much as they said they would and using the certainty response to adjust the bids (Champ, Bishop, Brown, & McCollum, 1997; Li & Mattsson, 1995), and (3) asking respondents before they give their hypothetical bid to consider their budget constraints (Loomis, Gonzalez-Caban, & Gregory, 1994; Neill et al., 1994). Although sometimes successful, each of these approaches has problems as well. Calibration is quite commodity-specific (Fox et al., 1998); the certainty scale does not provide an obvious cutoff point, reducing its accuracy; and the budgetary reminder has been largely ineffective.

This paper tests an approach similar to the budgetary reminder, in that it attempts to get respondents to avoid over-estimations, but it employs a much more direct attack on hypothetical bias. It instructs people, when answering a contingent willingness to pay question, to respond as if they were *really* spending their money. At least three studies, described next, have used this approach. The latter two show that the approach can achieve considerable success.

---

[2] Not all comparisons have found an over-estimation. For example, in an open-ended mail survey requesting donations to a land preservation fund, MacMillan et al. (1999) found no over-estimation when they adopted a campaign-style format, similar to typical solicitations for donations, for the contingent and real cash instruments. See also Dickie, Fisher, and Gerking (1987)

Past Attempts

Studies of attempts to talk people out of their hypothetical bias have compared hypothetical and real payment treatments with a hypothetical treatment employing a script instructing respondents to answer as if they were really spending their money. In the first study, Loomis et al. (1996) obtained sealed bids for an art print. In the "hypothetical with reminder" treatment, participants were encouraged to think about what they would "honestly" be prepared to pay if the auction were for real. In the two-paragraph entreaty, they were reminded of their budget constraint but also told, among other things, "Although the question is hypothetical, we want you to answer as if it were real—as if you were participating in a real sealed-bid auction and would really have to pay your dollar amount if you were the highest bidder" (p. 453). Mean bids from separate samples were $42 in the hypothetical treatment, $26 in the hypothetical-with-reminder treatment, and $14 in the actual payment treatment. The reminder helped reduce hypothetical bias, but was insufficient to remove it.[3]

In the second study, Cummings and Taylor (1999) used a much more thorough script, which they called "cheap talk," in their effort to remove bias from a hypothetical referendum on contributing $10 to fund a public good. In the eight-paragraph script, the authors not only encouraged participants to act as if the referendum were for real, but also explicitly addressed the problem of hypothetical bias—of how people in hypothetical bidding situations tend to over-estimate their willingness to pay—and discussed possible reasons for the phenomenon. Their approach met with notable success. For two of their three goods—the donation to the Nature Conservancy for Costa Rican rain forest protection mentioned above and a separate Nature Conservancy donation to protect land in the state of Georgia—the percentages of participants voting to donate $10 were nearly identical in the real payment and hypothetical-with-cheap-talk treatments, and these percentages were significantly below those of the standard hypothetical treatment. And for the other good—the Albuquerque citizens' guide—the percent voting yes when the cheap talk script was used was within 5 percentage points of its real referendum counterpart. This study is important not only for its demonstration of a promising approach for

[3] Cummings and Taylor (1999) cite two additional working papers that unsuccessfully used similarly short scripts to remove hypothetical bias.

adjusting for hypothetical bias, but also because it used a referendum format, which is arguably the most appropriate contingent valuation format for valuing public goods (Hoehn & Randall, 1987).

Finally, List (2001) used the Cummings and Taylor cheap talk script in a second-price auction to purchase a baseball card. Respondents from two groups of card collectors (professional card dealers and nondealers) were interviewed individually at a sports card show and given an opportunity to bid (either hypothetically or for real) on a card with a book value of about $200. For nondealers, the real payment and hypothetical-with-cheap-talk treatments each produced a mean bid of $26, compared with a mean bid of $49 in the hypothetical treatment. However, the professional dealers were largely unaffected by the cheap talk script, bidding a mean of $116 in the hypothetical treatment and $108 in the cheap talk treatment, compared with a mean of $60 in the actual auction. As in the Cummings and Taylor study, the cheap talk script worked remarkably well when bidders were unfamiliar with bidding for the good at issue. The lack of success with professional dealers raises the question: Are there other circumstances in which the script does not correct effectively for hypothetical bias?

The study reported in this discussion paper further tested the Cummings and Taylor script. Our bidding format, subject pool, recruitment enticement, and good were similar to theirs, in that students were paid $10 and participated in actual or hypothetical referenda on donating to a public good. We extended their work in two ways. First, whereas they used only one donation amount, $10, we varied the amount across referenda to investigate the effect of the amount on the success of the cheap talk script. The practical reason for such a test is that, if the script is to work in dichotomous choice contingent valuation, it must work at a range of bid levels. The more fundamental objective was to shed some light on when and how the script corrects for hypothetical bias. A side benefit of this study is that we tested for hypothetical bias in referenda at several payment amounts, the first such test of which we are aware. Second, in a further attempt to understand the workings of the script, we tested the effect of anticipating that the vote will be made public on hypothetical bias and on the success of the script.

We accept as given that when placed in a real payment situation, each participant will forms an estimate of willingness to pay (or casts a vote reflecting that willingness to

pay). The question is, in a hypothetical situation, will the cheap talk script lead the participant to the response that would be forthcoming in a real payment situation, or will it create a separate response that only equals the real payment response under certain circumstances? We suspected that the latter possibility was more likely, and specifically that the script—which as seen below is a very direct and lengthy attack on hypothetical bias—would perform best when hypothetical bias was greatest. Our hunch was that a script that worked well when the bias was substantial would be ineffective or over-correct when the bias was smaller.

Methods

In groups, students at the University of Massachusetts participated in a session of three or more stages, for which they were each paid $10 in cash at the outset. The referendum stage(s) of the sessions were preceded by an unrelated task and followed by a related but quite different task, a survey intended to measure tendency to respond in a socially desirable manner.

Each of 491 subjects participated in one of 16 different experimental versions. Each version included one or two referenda. The initial 12 versions included one secret ballot referendum; these 12 different referenda were created by crossing three levels of "reality" of payment (hypothetical, hypothetical with cheap talk script, and real payment) with four payment levels ($1, $3, $5, and $8). The next three versions were identical to the initial three $8 payment referenda except that participants expected that they would have to discuss their votes with the group. A final version included two secret ballot referenda at the $8 payment level, a hypothetical referendum followed by a real payment referendum.

The hypothetical with cheap talk (hereafter just cheap talk) versions of the referenda differed from the comparable hypothetical versions only in that the former included an extensive entreaty to respond as if the vote were for real. Sample sizes for the 12 initial versions are listed in Table 1, and those of the open referenda are listed in Table 5. There were 34 participants in the dual referenda version, who are included among the 52 participants listed in Table 1 for the $8 hypothetical referendum.

Table 1. Sample size of initial 12 referenda

	Payment amount			
	$1	$3	$5	$8
Hypothetical	31	37	33	52
Cheap talk	33	43	31	52
Real	28	37	30	17
Total	92	117	94	121

The referendum stage(s) of a session began with a description of and justification for the proposition to be voted on. It was described as an effort to make additional scholarship funds available to deserving students wishing to attend the university (Figure A1, Appendix A). The proposition ($8 version) was then stated as follows:

> "Everyone here in this room will contribute $8.00 to the *University of Massachusetts – Amherst Scholarship Fund.* The contribution is to be used to provide scholarship aid to needy and deserving students at the University of Massachusetts – Amherst."

The potential outcomes of the voting procedure were then explained, to the effect that if more than 50% of the participants in the room voted yes, all participants would donate the stated amount, and otherwise no one would pay. Figure A2 (Appendix A) shows the wording of the explanation in the hypothetical versions. The real and hypothetical versions differed only in the tense used to describe the outcomes. For example, the first sentence of the real version read "If more than 50% of you vote YES on this proposition, all of you pay $8.00 – I will collect $8.00 from each of you…"; whereas the first sentence of the hypothetical version read "If more than 50% of you were to vote YES on this proposition, all of you would pay $8.00 – I would collect $8.00 from each of you…".

At this point, in the cheap talk versions a lengthy discussion of hypothetical bias was presented (Figure A3, Appendix A). The script we used is identical to that used by Cummings and Taylor except for differences related to changes in the good. In this script, participants are told of the problem of hypothetical referenda—that when money payments are at issue people tend to vote differently in a hypothetical referendum than they do in a real referendum. Results of past studies are presented showing a higher percentage voting yes in a hypothetical referendum than in a real referendum. The reasons for this difference are then discussed, including the notion that in a referendum to do something good for

others, people tend to focus on doing good and tend to ignore the cost to them when it is hypothetical, but they temper that inclination with considerations of other options for their money when the referendum is for real. The cheap talk script ended with the following request: "I ask you to vote just exactly as you would vote if you were really going to face the consequences of your vote: which is to pay money if the proposition passes."

The vote was recorded on individual printed ballots that were placed in a ballot box. Before the vote, participants were reminded that their $10 participation fee was their money, which they earned by spending their time helping a research project (the cheap talk version of this reminder is included near the end of Figure A3). In addition, respondents in the open versions were told:

> "Most referenda are conducted as part of general elections where people tend to discuss their votes with friends, coworkers, and acquaintances. Therefore, to make our referendum comparable to other referenda, you will discuss your vote with the other people in this room. After you have cast your votes and handed in your ballots, we will ask each of you to state how you voted and explain *why* you voted the way you did."

Participants in the secret versions received no such instruction.

Results

As will be explained in more detail below, percentages voting yes varied widely among the referenda, from a low of 21 percent to a high of 82 percent.

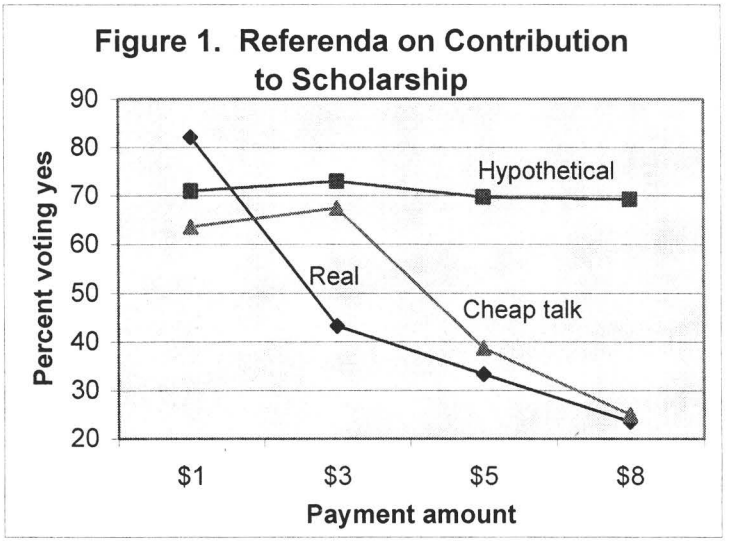
Table 2. Percent voting yes in initial 12 referenda

	Payment amount			
	$1	$3	$5	$8
Hypothetical	71	73	70	69
Cheap talk	64	67	39	25
Real	82	43	33	24

Effect of payment amount. Results of voting in the initial 12 referenda are presented in Figure 1. Voting in the four hypothetical referendum was insensitive to payment amount, with 69% to 73% of participants voting yes (Table 2).[4] At the 95%

[4] To test whether a yes vote of about 70% was an invariable result for public goods with our subject population, we also ran a hypothetical referendum (with 19 subjects) on donating $8 to a different scholarship fund, one usable at universities throughout the U.S., and observed a 47% yes vote. This extra

confidence level we cannot reject the hypothesis of no difference in voting among the four referenda (Table 3).[5] However, in the real referenda, percent voting yes decreased monotonically as payment amount increased, from 82% at the $1 level to 24% at the $8 level (percent voting yes is significantly greater at the $1 payment level than at the other three payment levels, Table 3). Clearly participants were sensitive to cost when spending real money. In the cheap talk referenda, percent voting yes decreased with payment amount across the $3, $5, and $8 amounts (with the $3 percentage yes being significantly greater than the $5 and $8 percentages yes, Table 3), but the $1 payment percentage yes was slightly below the $3 payment percentage yes.



Examining the tendency toward hypothetical bias, percentages voting yes in the hypothetical referenda were significantly greater than in the real referenda at the $3, $5, and $8 levels (Table 4). At the $1 level, the percentage voting yes was actually greater—though not significantly so—in the real referendum than in the hypothetical referendum, a result that may reflect the lack of salience of a $1 payment.

Table 3. Tests of equal proportions voting yes by payment amount (Fisher's exact p-value, two-sided test)

Null hypothesis	Hypothetical	Cheap talk	Real
$1 = $3	1.000	0.809	0.002
$1 = $5	1.000	0.079	0.000
$1 = $8	1.000	0.001	0.000
$3 = $5	0.796	0.019	0.458
$3 = $8	0.814	0.000	0.229
$5 = $8	1.000	0.221	0.529

referendum suggests sensitivity to the nature of the good in the hypothetical condition.

[5] Two-sided tests of hypotheses are reported herein using Fisher's exact test. One-sided tests, of course, yield lower probability levels, but reach the same conclusions for our data at the 0.05 significance level, except for the comparison of $1 and $5 in the cheap talk treatment, for which the probability level drops from 0.079 with a two-sided test to 0.040 with a one-sided test.

Table 4. Tests of equal proportions voting yes by treatment (Fisher's exact p-value, two-sided test)

Null hypothesis	Payment amount			
	$1	$3	$5	$8
Real = hypothetical	0.370	0.018	0.006	0.001
Real = cheap talk	0.154	0.042	0.791	1.000
Hypothetical = cheap talk	0.601	0.632	0.023	0.000

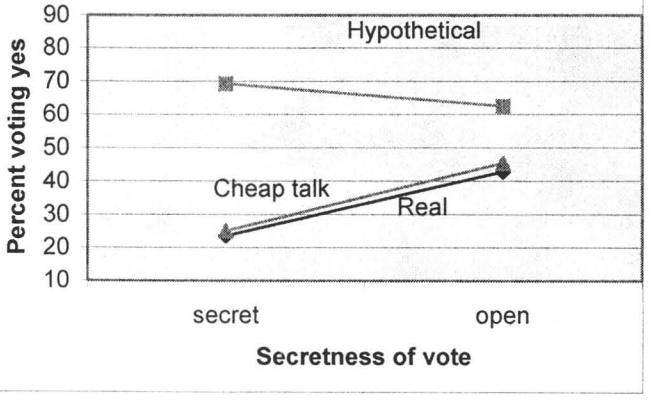
The cheap talk script had the desired effect of eliminating the hypothetical bias at the higher payment levels, dropping percent voting yes to within one percentage point of what it was in the real referendum at the $8 level, and to within 6 percentage points at the $5 level. At both of these payment amounts, the hypothetical percentages yes are significantly greater than the corresponding cheap talk percentages (Table 4). But at the $3 payment level the cheap talk script had a very modest effect, producing a percentage yes only 6 points below the hypothetical percentage yes, and significantly greater than the real percentage yes (Table 4). And at the $1 level, no correction was needed but a correction nevertheless occurred, with the cheap talk percentage yes being 18 percentage points below that of the real referendum, though the difference is not statistically significant.

Effect of openness of the vote at the $8 payment level. As seen in Figure 2, knowing that one had to reveal and explain one's vote to others produced a lower percent voting yes in the hypothetical referendum (63%, Table 5) than in the comparable secret ballot referendum (69%, Table 2), and a higher percent yes in the real referendum (43%) than in the comparable secret ballot referendum (24%), narrowing the difference in percent yes between the hypothetical and real payment versions. Neither of these changes in percent voting yes are significant (p = 0.606 for the

Table 5. Effect of anticipating open discussion of one's vote, $8 payment amount

	Sample size	Percent voting yes
Hypothetical	24	63
Cheap talk	22	45
Real	21	43

Figure 2. Referenda on Contribution to Scholarship: $8 payment



hypothetical referenda and p = 0.307 for the real referenda, Fisher's exact two-sided tests), but the changes do suggest the possibility that in the hypothetical referendum the prospect of explaining one's vote leads some people to more carefully consider their true willingness to pay, and that in the real referendum some people value appearing to others as generous.

More importantly for our interest here, the cheap talk script again worked well, dropping hypothetical percent yes to within two percentage points of what it was in the real referendum. Although neither the real nor the cheap talk percentage yes is significantly lower than the hypothetical percentage (p = 0.238 and p = 0.375, respectively, Fisher's exact two-sided tests), this result for open referenda at least does not detract from the earlier finding of effectiveness of the script at the $8 payment level.

Effect of a prior hypothetical referendum on a real referendum, secret ballot at the $8 payment level. Having established that, at the $8 payment level, the cheap talk script yields voting in a hypothetical referendum that mimics that of a real referendum, we wondered whether having voted yes in an ordinary (without cheap talk) hypothetical referendum would raise percent yes in a subsequent real referendum to what it was in the prior hypothetical referendum. That is, we wondered whether voting in a real referendum was sensitive to prior hypothetical responses. To address this we held a version that included both referenda, allowing a within-subject comparison.

The hypothetical referendum results are very similar to those of the comparable solo hypothetical referendum, with 68% and 72%, respectively, voting yes (the combined percent yes is 69%, Table 2). However, in the real payment referendum that followed the hypothetical referendum 44% voted yes, which is substantially above the 24% voting yes in the solo real payment referendum, though the difference is not significant (p = 0.222, Fisher's exact two-sided test).[6] Neither is the 44% voting yes in the real referendum significantly different from the 68% in the prior hypothetical referendum (p = 0.087, Fisher's exact two-sided test). Thus, while we have a suggestion that a hypothetical referendum can increase the likelihood of a yes vote in a subsequent real referendum, the evidence is weak.

---

[6] At the individual subject level, 9 of 23 subjects (39%) switched from a yes vote in the hypothetical referendum to a no vote in the subsequent real referendum.

Discussion

This study joins a number of prior studies demonstrating hypothetical bias in elicitations of willingness to pay by comparing responses from hypothetical and real payment conditions. Two of these studies (Cummings et al., 1997; Cummings & Taylor, 1999) used referenda about public goods, but employed only a $10 payment amount. We extend prior work by testing for bias in several secret ballot referenda differing by payment amount. Among the four payment amounts ($1, $3, $5 and $8), the percent voting in favor in the hypothetical referenda remained nearly constant, whereas when the referendum was for real the percent in favor steadily dropped as the amount was raised, causing hypothetical bias to increase with payment amount.

The sensitivity of the vote to payment amount in the real referenda is expected,[7] but the insensitivity to payment amount in the hypothetical referenda is not, as previous studies of hypothetical responses have found that proportion yes drops as bid level increases. Such sensitivity has been found for a private good at bid levels ($1 and $5) similar to those used here (Blumenschein, Johannesson, Blomquist, Liljas, & O'Connor, 1998). And dichotomous choice contingent valuation studies of public goods regularly show a significant negative relation between stated willingness to pay and bid level (for contingent referenda examples, see Carson et al., 1998; Loomis et al., 1994). However, unlike our study, most contingent valuation studies use a wide range of payment amounts. Payment amount in our study was limited by the $10 participation fee. Thus, the lack of sensitivity to payment amount that we found cannot be generalized to higher payment amounts or to bids for private goods.

Why was the hypothetical vote invariant to payment amount, thereby causing the substantial hypothetical bias we observed? Four possibilities come to mind, the first three of which are suggested in the cheap talk script. First, some participants in hypothetical referenda may try to estimate their real willingness to pay but may over-estimate. Second, some may ignore their budget constraint, thereby responding as if they faced none. Third, some may use their vote to indicate whether or not they think the cause is worthy of

---

[7] In addition to the obvious effect of rising payment amount on willingness to vote yes—that being the personal opportunity cost of making the payment to fund the scholarship—voters may also be sensitive to payment amount because the do not want to force others to pay. Such other-focused preferences would probably not arise in the hypothetical referenda.

support. Fourth, some may use their vote to indicate whether or not they think such a referendum should be held for real. In essence, participants acting according to the latter three of these possibilities are answering different questions from the one they were asked in the hypothetical referendum.

The Cummings and Taylor cheap talk script worked remarkably well at the two larger payment amounts, lowering the percent voting in favor to within 2 ($8) and 6 ($5) percentage points of what it was in the real referenda. This result is similar to those obtained by Cummings and Taylor in their public good referenda, which used a $10 payment with a similar subject pool. In addition to these secret referenda, we also held referenda at the $8 payment level when participants knew they had to reveal how they had voted (the open condition). Hypothetical bias was lower in this condition than in the $8 secret condition, but still substantial, and the cheap talk script once again lowered the percent voting in favor to within 2 percentage points of what it was in the real referendum, further enhancing confidence in the robustness of the script at the higher payment level.

However, the script significantly under-corrected at the $3 payment level, lowering percentage voting in favor only 6 percentage points below that of the hypothetical condition. One possible explanation for the script's ineffectiveness at the $3 payment level is that some participants may have thought it implausible that they would vote against the referendum at such a low payment. That is, they may not have envisioned significant bias at that level.

Finally, at the $1 level the script caused a 7-point drop in percentage yes, nearly the same correction as occurred at the $3 payment level. However, in this case a correction was not needed, as percentage yes was higher in the real than in the hypothetical referendum.[8] In any case, the percentages yes of the three $1 referenda were not significantly different from each other.

In general, when used with public good referenda, the cheap talk script appears to increase in effectiveness as payment level, and correspondingly hypothetical bias, increase—though the evidence of this is limited to the relatively modest payment levels used here and by Cummings and Taylor. If this trend holds at higher bid levels, it bodes

[8] Cummings and Taylor (1999) tested for over-correction in a referendum where the percentage voting yes was also slightly but not significantly higher in the real referendum than in the hypothetical referendum, and found no over-correction with the cheap talk script, but their test was at the $10 payment level.

well for contingent valuation, because hypothetical bias in dichotomous choice contingent valuation of public goods tends to remain substantial at higher payment levels (see Brown et al., 1996 for supporting data).

Turning now to the dual referenda version, our result of no significant effect of the prior hypothetical referendum on voting in the subsequent real payment referendum is in concert with what several other studies found for private goods. Hypothetical willingness to pay responses had no aggregate effect on subsequent real cash responses in dichotomous choice bidding for common consumer items (Cummings et al., 1995), in open-ended bidding in an auction for an art print (Loomis et al., 1996); and at one of two payment levels in dichotomous choice bids for a consumer good (Blumenschein et al., 1998). Our result is tentative because the difference in voting, though not significant, was substantial. A larger sample size might produce a significant difference. But based on the available evidence, we are unable to conclude that votes in a referendum are any different from dichotomous choice responses about private goods regarding the effect of hypothetical bidding on subsequent real cash bidding. If the finding holds, it offers an opportunity for exploring the reasons for hypothetical bias with individual respondents (by, or example, asking respondents to explain the difference).

Understanding how payment amount or other factors alter the effectiveness of the cheap talk script will require a better understanding of its workings at the individual level, for people may differ in both the correction required and the effect of the script. An understanding at the individual level could show to what extent the cheap talk script causes all respondents to correct by an equal amount regardless of their level of bias (thus causing over-correction in some people and under-correction in others), versus causing each individual to remove the bias whatever its magnitude. Some insight into this issue is offered by further examination of the data collected by Loomis et al. (1996) in a study, mentioned above, that used a short version of cheap talk, called "reminder," with a first-price auction for an art print. That examination, presented in Appendix B, shows that respondents largely fell into two groups, those who removed their bias and those who remained unaffected by the script. Over-correction was almost non-existent. We cannot know how the effect of the cheap talk script on individual bias differs from the effect of the milder reminder used in the art print study. The differences in script and in response mode

(referendum versus first price auction) confound conclusive comparisons. But the art print data at least does not dampen hopes that the cheap talk script works as intended, by individually removing bias without causing over-correction.

The Cummings and Taylor script is long, and would thus be difficult to use on the phone or in the mail. A much shorter version of the entreaty mentioned above (used by Loomis et al., 1996), which did not specifically educate respondents about hypothetical bias, helped reduce hypothetical bias but under-corrected on average. Although we lack a definitive test of the effect of the difference between the Loomis et al. and the Cummings and Taylor scripts, it appears that the entreaty has a greater impact if it not only instructs participants to act as if the response were for real, but also explains that responses to a hypothetical question are often very different from those to a real payment question. Only further research will determine whether the script can be shortened without loss of impact.

If further tests continue to show that the cheap talk script helps remove bias at all but the trivially small (e.g., $1) payment amounts, and works extremely well at higher payment amounts, it would be unduly prudent to delay its application to dichotomous choice contingent valuation—especially if additional research shows that the script can be shortened in length without loss of effect.

Literature Cited

Blackburn, M., Harrison, G. W., & Rutstrom, E. E. (1994). Statistical Bias Functions and Informative Hypothetical Surveys. *American Journal of Agricultural Economics, 76*(5), 1084-1088.

Blumenschein, K., Johannesson, M., Blomquist, G. C., Liljas, B., & O'Connor, R. M. (1998). Experimental results on expressed certainty and hypothetical bias in contingent valuation. *Southern Economic Journal, 65*(1), 169-177.

Boyce, R. R., Brown, T. C., McClelland, G. D., Peterson, G. L., & Schulze, W. D. (1989). Experimental evidence of existence value in payment and compensation contexts. In K. J. Boyle & T. Heekin (Eds.), *Western Regional Research Project W-133: Benefits and Costs in Natural Resources Planning* (Interim Report 2 ed., pp. 305-336). Orono, ME: University of Maine.

Brown, T. C., Champ, P. A., Bishop, R. C., & McCollum, D. W. (1996). Which response format reveals the truth about donations to a public good? *Land Economics, 72*(2), 152-166.

Carson, R. T., Hanemann, W. M., Kopp, R. J., Krosnick, J. A., Mitchell, R. C., Presser, S., Ruud, P. A., Smith, V. K., Conaway, M., & Martin, K. (1998). Referendum design and contingent valuation: The NOAA panel's no-vote recommendation. *Review of Economics and Statistics, 80*(3), 484-487.

Champ, P. A., Bishop, R. C., Brown, T. C., & McCollum, D. W. (1997). Using donation mechanisms to value nonuse benefits from public goods. *Journal of Environmental Economics and Management, 33*, 151-162.

Cummings, R. G., Elliott, S., Harrison, G. W., & Murphy, J. (1997). Are hypothetical referenda incentive compatible? *Journal of Political Economy, 105*(3), 609-621.

Cummings, R. G., Harrison, G. W., & Rutstrom, E. E. (1995). Homegrown values and hypothetical surveys: Is the dichotomous choice approach incentive-compatible? *American Economic Review, 85*(1), 260-266.

Cummings, R. G., & Taylor, L. O. (1999). Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method. *American Economic Review, 89*(3), 649-665.

Dickie, M., Fisher, A., & Gerking, S. (1987). Market transactions and hypothetical demand: A comparative study. *Journal of the American Statistical Association, 82*(397), 69-75.

Fox, J. A., Shogren, J. F., Hayes, D. J., & Kliebenstein, J. B. (1998). CVM-X: Calibrating Contingent Values with Experimental Auction Markets. *American Journal of Agricultural Economics, 80*(3), 455-465.

Hoehn, J. P., & Randall, A. (1987). A satisfactory benefit cost indicator from contingent valuation. *Journal of Environmental Economics and Management, 14*, 226-247.

Kealy, M. J., Dovidio, J. F., & Rockel, M. L. (1988). Accuracy in Valuation is a Matter of Degree. *Land Economics, 64*(2), 158-171.

Li, C.-Z., & Mattsson, L. (1995). Discrete Choice under Preference Uncertainty: An Improved Structural Model for Contingent Valuation. *Journal of Environmental Economics and Management, 28*, 256-269.

List, J. A. (2001). Do explicit warnings eliminate the hypothetical bias in elicitation procedures? Evidence from field auctions for sportscards. *American Economic Review, 91*(5), 1498-1507.

Loomis, J., Brown, T., Lucero, B., & Peterson, G. (1996). Improving validity experiments of contingent valuation methods: Results of efforts to reduce the disparity of hypothetical and actual WTP. *Land Economics, 72*(4), 450-461.

Loomis, J., Brown, T. C., Lucero, B., & Peterson, G. (1997). Evaluating the validity of the dichotomous choice question format in contingent valuation. *Environmental and Resource Economics, 10*, 109-123.

Loomis, J., Gonzalez-Caban, A., & Gregory, R. (1994). Do reminders of substitutes and budget constraints influence contingent valuation estimates? *Land Economics, 70*(4), 499-506.

MacMillan, D. C., Smart, T. S., & Thorburn, A. P. (1999). A Field Experiment Involving Cash and Hypothetical Charitable Donations. *Environmental and Resource Economics, 14*, 399-412.

Neill, H., Cummings, R., Ganderton, P., Harrison, G., & McGuckin, T. (1994). Hypothetical surveys and real economic commitments. *Land Economics, 70*(2), 145-154.

Appendix A. Experimental Material

Figure A1. Description of Referendum ($8 payment version)

The University of Massachusetts at Amherst receives applications from a broad range of prospective students. Many of these applicants come from poor families who cannot afford the tuition, fees, room and board, and other expenses of attending the University. A 1998 survey conducted for the State Board of Higher Education found that as many as 23% of all high school graduates fail to go on to college, or drop out of college, for financial reasons. The State of Massachusetts is heavily dependent on an educated workforce for its high-tech industries and cannot afford to lose this potential source of qualified workers.

As a State University, UMass–Amherst is committed to providing accessible higher education to every qualified resident. Not only does this help to prepare a well trained work force, it is also an essential tool for social mobility and equity. To achieve these goals, UMass–Amherst tries to provide full or partial scholarships to as many needy and deserving students as possible. Although the University receives some State funds for scholarships, the amount available is woefully inadequate. Consequently, UMass–Amherst maintains a general scholarship fund and is actively soliciting donations to increase the number of scholarships it can offer.

If everyone in this room were to contribute $8.00, these moneys would be added to the UMass–Amherst Scholarship Fund and would be used to help deserving students attend the University.

Figure A2. Description of Voting Procedure Outcomes (hypothetical $8 payment version)

1. If more than 50% of you were to vote YES on this proposition, all of you would pay $8.00 – I would collect $8.00 from each of you – and we would give this money to the University of Massachusetts – Amherst with instructions that the money is to be credited to the University Scholarship Fund.

We wouldn't send cash. We would take your cash, write a check for the total amount of money, and the check would be mailed to the University Development Office which manages all donations. I would put the check in this stamped envelope addressed to the Development Office. <HOLD UP ENVELOPE> I would ask one of you to put the envelope in the mail box on the third floor. When I received a receipt for the money from the Development Office, I would make it available for your inspection on the bulletin board near Room 623 in Tobin Hall.

2. If 50% or fewer of you were to vote YES on this proposition, no one would pay $8.00, and we would not send a check to the UMass Scholarship Fund.

Figure A3. Cheap Talk Script ($8 payment version)

Before we have our vote, I want to talk to you about a problem that we have in studies like this one. As I told you a minute ago, this is a hypothetical referendum – not a real one. No one will actually pay money at the end of the vote. But I also asked you to respond to the vote as though the result of your vote could involve a real cash payment by you.

And that's the problem. In most studies of this kind, people seem to have a hard time doing this. They vote differently in a hypothetical referendum, where they don't really have to pay money, than they do in a real referendum, where they really could have to pay money. For example, in a recent study, several different groups of people voted on a referendum just like the one you are about to vote on. Payment was hypothetical for these groups, as it will be for you. No one had to pay money if the referendum passed. <SHOW OVERHEAD OF SURVEY RESULTS> The results of these studies were that on average, across the groups, 45% of them voted YES. With another set of groups with similar people voting on the same referendum as you will vote on here, but where payment was real and people really did have to pay money if the referendum passed, the results on average, across the groups were that 27% voted YES. That's quite a difference, isn't it?

We call this a "hypothetical bias". "Hypothetical bias" is the difference that we continually see in the way people respond to hypothetical referenda as compared to real referenda. In the real referendum, where people knew they would have to pay money if the referendum passed, 27% voted yes and 73% voted no. When payment was hypothetical and people knew they would not pay anything if the referendum passed, just like your vote today, 45% voted yes and 55% voted no.

How can we get people to think about their vote in a hypothetical referendum like they think in a real referendum, where if enough people vote YES, they'll really have to pay money? How do we get them to think about what it means to really dig into their pocket and pay money, if in fact they really aren't going to have to do it?

Let me tell you why I think that we continually see this hypothetical bias, why people behave differently in a hypothetical referendum than they do when the referendum is real. I think that when we hear about a referendum that involves doing something that is basically good – helping people, improving environmental quality, or anything else – our basic reaction in a hypothetical situation is to think: sure, I would do this. I really would vote YES to spend the money – I really, really, think I would. What our YES vote really means in this case is that we are basically good people, and that we would like to see good things happen.

But when the referendum is real, and we would actually have to spend our money if it passes, we think a different way. We basically still would like to see good things happen, but when we are faced with the possibility of having to spend money, we think about our options: if I spend money on this, that's money I don't have to spend on other things. If I spend money to help needy students, that's money I don't have to spend on groceries, go to a movie, or perhaps give to some environmental organization. So when the payment is real if the referendum passes, we vote in a way that takes into account the limited amount of money we have. We vote realizing that we just don't have enough money to do everything we might like to do. This is just an opinion, of course, but it's what may be going on in hypothetical referenda.

In any case, the only way that we know to get people like you to vote in our hypothetical referendum just like you'd vote if the referendum was real is to simply ask you: in the vote that we're going to take in a few minutes, please do the following:

- Think about what you're voting on. If this were real—if more than 50% of you voted YES, you would actually have to dig into your pocket and pay $8.00 right now—do you really want to support the UMass-Amherst Scholarship Fund enough that you would be willing to spend the money?
- Also, let me make clear that the $10.00 participation fee that you were paid today is your money. You've spent your time helping us in our research, and you've earned it! You were told that the money is yours, believe it! So, if I were in your shoes, and I was asked to vote YES or NO on this proposition that requires all of us to pay $8.00, I would think about how I feel about spending my money this way. When I got ready to vote, I would ask myself: if this were a real referendum, and I had to pay $8.00 if the referendum passed, do I really want to spend my money this way. If I really did, I would vote YES; if I didn't, I would vote NO – I wouldn't throw my money around. That's just my opinion, of course. You must do whatever you want to do.
- In any case, I ask you to vote just exactly as you would vote if you were really going to face the consequences of your vote: which is to pay money if the proposition passes.

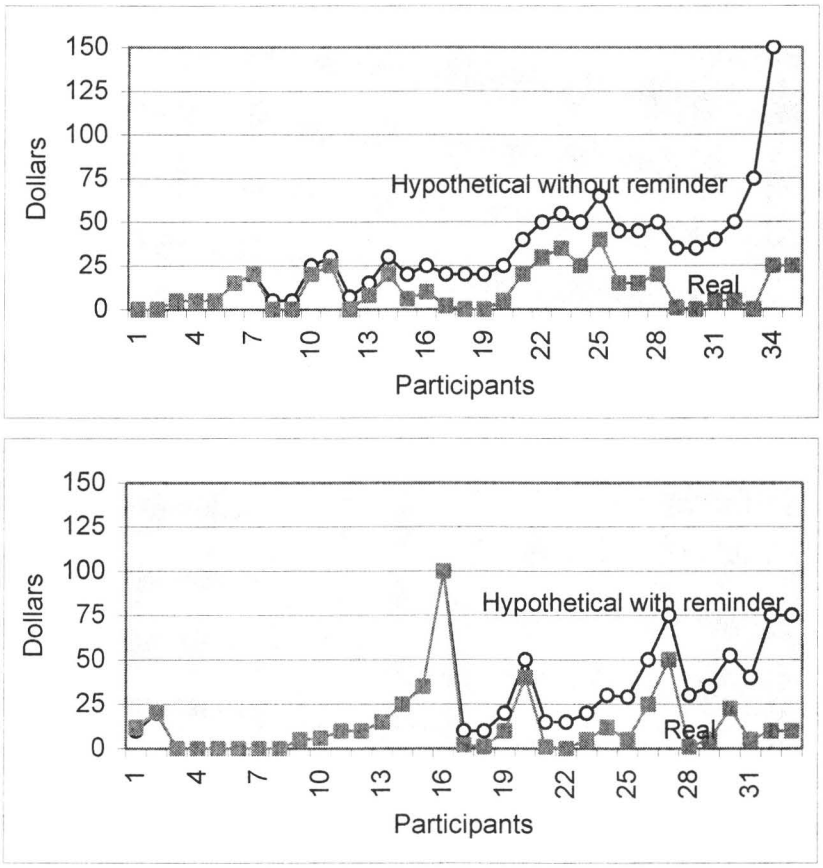
Appendix B. Further Evidence on Effect of Entreaties to Avoid Hypothetical Bias

In the art print study (Loomis et al., 1996), the hypothetical auctions, one with a "reminder" and another without, were followed by real payment auctions. The reminder encouraged participants to bid as if the auction were for real. Participants in these two hypothetical auctions did not know that a real payment auction would follow. A third group participated only in a real payment auction. Sample sizes of the three groups varied from 32 to 35 participants. Mean bids of the three real payment auctions varied from $12 to $14, suggesting that the prior hypothetical auctions held with two of the groups did not affect bidding in the subsequent real payment auctions. As previously mentioned, the mean hypothetical bids were $42 without the reminder and $26 with the reminder. Of particular interest here is the distribution of hypothetical bias (i.e., the difference between hypothetical and real payment bid) across participants in the two groups that bid in both kinds of auctions (Figure B1).

Figure B1 shows three findings of interest. First, bias was less common with the reminder. Eighty percent of participants in the group without the reminder exhibited a bias, whereas only 52% in the group with the reminder had a positive bias. Second, over-correction with the reminder was minimal. Only two participants (numbers 1 and 2 in the bottom graph of Figure B1) bid more in the real payment auction than in the hypothetical auction, and the differences were within $2. Third, among participants with a positive bias, the average magnitude of the bias did not differ much between treatments. Ignoring a $400 bid in the without-reminder treatment (which does not show in the upper right corner of the top graph), the mean positive bias was $26 without the reminder and $25 with the reminder (the medians are $20 and $24, respectively).

Further evidence is offered by an additional treatment in the Loomis et al. (1996) study, one identical to the hypothetical with-reminder treatment of Figure B1 except that a dichotomous choice response preceded the two open-ended responses of interest here (see Loomis, Brown, Lucero, & Peterson, 1997). In this treatment of 54 participants, mean open-ended willingness to pay was $23 in the hypothetical auction and $12 in the real auction, only one with-reminder participant over-corrected (by $1), 44 percent of the participants had a positive bias, and positive bias averaged $24. These results are very close to those of the with-reminder treatment shown in Figure B1.

Figure B1. Open-ended Bids for an Art Print, with Participants Sorted in Order of Increasing Bias



Thus, the principal effect of the reminder appears to have been to decrease the number of participants with a bias. With the exception of the possible effect on extreme hypothetical bids (which involves only one participant in the without-reminder treatment), the reminder left the average positive bias unchanged, suggesting that the size of participants' hypothetical bias played no role in their susceptibility to the reminder.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
ACADEMIC PRESS
Journal of Environmental Economics and Management 46 (2003) 353–361
JOURNAL OF ENVIRONMENTAL ECONOMICS AND MANAGEMENT
http://www.elsevier.com/locate/jeem

Further tests of entreaties to avoid hypothetical bias in referendum contingent valuation

Thomas C. Brown,[a,*] Icek Ajzen,[b] and Daniel Hrubes[c]

[a] *Rocky Mountain Research Station, U.S. Forest Service, Fort Collins, CO 80526, USA*
[b] *Department of Psychology, University of Massachusetts, Amherst, MA 01003, USA*
[c] *Department of Psychology, College of Mount Saint Vincent, Riverdale, NY 10471, USA*

Received 16 February 2001; revised 16 October 2001

Abstract

Over-estimation of willingness to pay in contingent markets has been attributed largely to hypothetical bias. One promising approach for avoiding hypothetical bias is to tell respondents enough about such bias that they self-correct for it. A script designed for this purpose by Cummings and Taylor was used in hypothetical referenda that differed in payment amount. In comparisons with behavior observed in otherwise identical real payment referenda, the script worked remarkably well at higher payment levels (dropping the likelihood of a yes vote to the level obtained in separate sample real referenda), but less well at a lower payment level. The repeated success of the script at higher payment amounts is a clear sign that additional research is warranted—research that will bring us a more complete understanding of how the entreaty works. This research may need to explore its workings at the individual respondent level.
© 2003 Elsevier Science (USA). All rights reserved.

JEL classification: Q20; Q26

Keywords: Hypothetical bias; Contingent valuation; Public goods; Referenda

1. Introduction

Given the goal of efficient management of environmental resources but the difficulty of estimating some benefits of environmental programs using revealed preference methods, hopes have been high that properly designed contingent markets would emulate actual markets. However, evidence has mounted that willingness to pay measured in contingent markets

*Corresponding author.
E-mail address: tcbrown@lamar.colostate.edu (T.C. Brown).

0095-0696/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0095-0696(02)00041-4

over-estimates payment observed in otherwise identical real cash markets. Though not a uniform result (e.g., [7,8]), differences between real cash (hereafter just real) and hypothetical willingness to pay, which we will call hypothetical bias, have typically been found both for private [2,6,11–13,15] and public [3,5,7] goods. These studies used a mixture of bidding mechanisms, suggesting that the over-estimation cannot be ignored.

One approach to avoiding hypothetical bias in contingent valuation is to ask respondents before they give their hypothetical bid to carefully consider their answer. An early version of this was to remind respondents of their budget constraint, but this approach proved to be largely ineffective [14,15]. This paper tests an approach similar to the budgetary reminder, in that it attempts to get respondents to avoid over-estimations, but it employs a much more direct attack on hypothetical bias. It instructs people, when answering a contingent willingness to pay question, to respond as if they were *really* spending their money. At least three studies, described next, have used this approach. These studies compared hypothetical and real payment treatments with a hypothetical treatment employing an entreaty urging respondents to answer as if they were really spending their money.

Loomis et al. [13] obtained sealed bids for an art print. In the two-paragraph entreaty used with the hypothetical treatment, respondents were reminded of their budget constraint but also were encouraged to think about what they would "honestly" be prepared to pay if the auction were for real and were told, among other things, "Although the question is hypothetical, we want you to answer as if it were real—as if you were participating in a real sealed-bid auction and would really have to pay your dollar amount if you were the highest bidder" (p. 453). Mean bids from separate samples were $42 in the hypothetical treatment, $26 in the hypothetical-with-entreaty treatment, and $14 in the actual payment treatment. The entreaty helped reduce hypothetical bias, but was insufficient to remove it.[1]

Cummings and Taylor [7] used a much more thorough script, which they called "cheap talk," in their effort to remove bias from a hypothetical referendum on contributing $10 to fund a public good. In the eight-paragraph script, the authors not only encouraged participants to act as if the referendum were for real, but also explicitly addressed the problem of hypothetical bias (of how people in hypothetical bidding situations tend to over-estimate their willingness to pay), presented data demonstrating hypothetical bias, and discussed possible reasons for the phenomenon. Their approach met with notable success. For two of their three goods—the donation to the Nature Conservancy for Costa Rican rain forest protection and a separate Nature Conservancy donation to protect land in the state of Georgia—the percentages of participants voting to donate $10 were nearly identical in the real payment and hypothetical-with-cheap-talk treatments, and these percentages were significantly below those of the standard hypothetical treatment. And for the other good—the Albuquerque citizens' guide—the percent voting yes when the cheap talk script was used was within 5 percentage points of its real referendum counterpart. This study is important not only for its demonstration of a promising approach for adjusting for hypothetical bias, but also because it used a referendum format, which is arguably the most appropriate contingent valuation format for valuing public goods [10].

[1] Cummings and Taylor [7] cite two working papers that also unsuccessfully used short scripts to remove hypothetical bias.

Recently, List [12] used the Cummings and Taylor cheap talk script in a second-price auction to purchase a baseball card. Respondents from two groups of card collectors (professional card dealers and nondealers) were interviewed individually at a sports card show and given an opportunity to bid (either hypothetically or for real) on a card with a book value of about $200. For nondealers, the real payment and hypothetical-with-cheap-talk treatments each produced a mean bid of $26, compared with a mean bid of $49 in the hypothetical treatment. However, the professional dealers were largely unaffected by the cheap talk script, bidding a mean of $116 in the hypothetical treatment and $108 in the cheap talk treatment, compared with a mean of $60 in the actual auction. As in the Cummings and Taylor study, the cheap talk script worked remarkably well when bidders were unfamiliar with bidding for the good at issue. The lack of success with professional dealers raises the question: Are there other circumstances in which the script does not correct effectively for hypothetical bias?

We further tested the Cummings and Taylor script. Our bidding format, subject pool, recruitment enticement, and good were similar to theirs, in that students were paid $10 and participated in actual or hypothetical referenda on donating to a worthy cause. But whereas Cummings and Taylor used only one donation amount ($10), we varied the amount across referenda to investigate the effect of payment amount on the success of the cheap talk script. The practical reason for such a test was to explore the script's utility for application to dichotomous choice contingent valuation, for if the script is to work in that context it must work at a range of bid levels. The more fundamental objective was to explore the script's limitations, for by finding situations where it does not work well—if indeed there are other such situations—and comparing those with situations where it does work, we can begin to form hypotheses about how it works.

2. Approach and assumptions

We accept as given that when placed in a real payment situation, each participant forms an estimate of willingness to pay (or casts a vote reflecting that willingness to pay). The question is, in a hypothetical situation, will the cheap talk script lead participants to the response that would be forthcoming in a real payment situation, or will it create a separate response that only equals the real payment response in certain circumstances? We suspected that the latter possibility was more likely, and specifically that the cheap talk script—a very direct and lengthy attack on hypothetical bias—would perform best when hypothetical bias was greatest. Our hunch was that a script that worked well when the bias was substantial would be less effective or over-correct when the bias was smaller. Loss of effectiveness is possible in situations of small bias if, for example, respondents do not think hypothetical bias is plausible. Over-correction is possible if the script alters responses when no alterations are needed.

One may question whether the referenda we and Cummings and Taylor used, which were about providing donations to worthy causes, are incentive compatible. A concern with referenda about pubic goods is that some voters may suspect that other options exist for providing the good, leading them to vote no when, in the absence of such suspicions, they would vote yes.[2]

[2] One alternative provision option is a future, more attractive referendum. Voters may suspect that other referenda could follow the current one [9]; indeed, in some states repeated referenda on the same funding issue are common [16].

Table 1
Sample size of 12 referenda

	Payment amount			
	$1	$3	$5	$8
Hypothetical	31	37	33	52
Cheap talk	33	43	31	52
Real	28	37	30	17
Total	92	117	94	121

We attempted to lessen the likelihood of such suspicions by informing participants of the great need for funds for the cause described (scholarships) and that whatever the participants donated would indeed be used, but we can make no claims of having avoided the potential concern. Rather, we take our experimental participants' votes in the real payment referenda on face value, as reflecting real willingness to pay in the context presented. We test whether the cheap talk script avoids hypothetical bias given whatever assumptions participants may have made about other options for providing the good. Because real referenda about public goods that completely avoid suspicions about other possible options for funding the good are probably rare, we suspect that our test results are relevant to many real-life referenda.

3. Methods

In groups, students at the University of Massachusetts participated in a session of three or more stages, for which they were each paid $10 in cash at the outset. The referendum stage of the sessions was preceded by an unrelated task and followed by a related but quite different task, a survey intended to measure tendency to respond in a socially desirable manner.

Each of 424 subjects participated in one of 12 different experimental versions including a secret ballot referendum. The 12 versions were created by crossing three levels of "reality" of payment (hypothetical, hypothetical with cheap talk script, and real payment) with four payment levels ($1, $3, $5, and $8). The hypothetical with cheap talk (hereafter just cheap talk) versions of the referenda differed from the comparable hypothetical versions only in that the former included an extensive entreaty to respond as if the vote were for real. Sample sizes for the 12 versions are listed in Table 1.

The referendum stage of a session began with a description of and justification for the proposition to be voted on. It was described as an effort to make additional scholarship funds available to deserving students wishing to attend the university.[3] The proposition ($8 version) was

[3] Complete texts of this and other scripts we used are available at the following website: http://www.fs.fed.us/rm/value/cheaptalkscript.html.

then stated as follows:

> *Everyone* here in this room will contribute $8.00 to the *University of Massachusetts—Amherst Scholarship Fund.* The contribution is to be used to provide scholarship aid to needy and deserving students at the University of Massachusetts—Amherst.

The potential outcomes of the voting procedure were then explained, to the effect that if more than 50% of the participants in the room voted yes, all participants would donate the stated amount, and otherwise no one would pay. The real and hypothetical versions differed only in the tense used to describe the outcomes. For example, the first sentence of the real version read "If more than 50% of you vote YES on this proposition, *all* of you pay $8.00—I will collect $8.00 from *each* of you..."; whereas the first sentence of the hypothetical version read "If more than 50% of you were to vote YES on this proposition, *all* of you would pay $8.00—I would collect $8.00 from *each* of you...". The process of collecting the money and writing and sending the check for the total amount was then explained.

At this point in the cheap talk versions a lengthy discussion of hypothetical bias was presented. The script we used is identical to that used by Cummings and Taylor except for differences related to changes in the good. In this script, participants are told of the problem of hypothetical referenda—that when money payments are at issue people tend to vote differently in a hypothetical referendum than they do in a real referendum. Results of past studies are presented showing a higher percentage voting yes in a hypothetical referendum than in a real referendum. The reasons for this difference are then discussed, including the notion that in a referendum to do something good for others, people tend to focus on doing good and tend to ignore the cost to them when it is hypothetical, but they temper that inclination with considerations of other options for their money when the referendum is for real. The cheap talk script ended with the following request: "I ask you to vote just exactly as you would vote if you were really going to face the consequences of your vote: which is to pay money if the proposition passes."

The vote was recorded on individual printed ballots that were placed in a ballot box. Before the vote, participants were reminded that their $10 participation fee was their money, which they earned by spending their time helping a research project.

4. Results

Percentages voting yes varied widely among the referenda, from a low of 21 percent to a high of 82 percent (Table 2, Fig. 1). Voting in the four hypothetical referenda was insensitive to payment

Table 2
Percent voting yes in 12 referenda

	Payment amount			
	$1	$3	$5	$8
Hypothetical	71	73	70	69
Cheap talk	64	67	39	25
Real	82	43	33	24

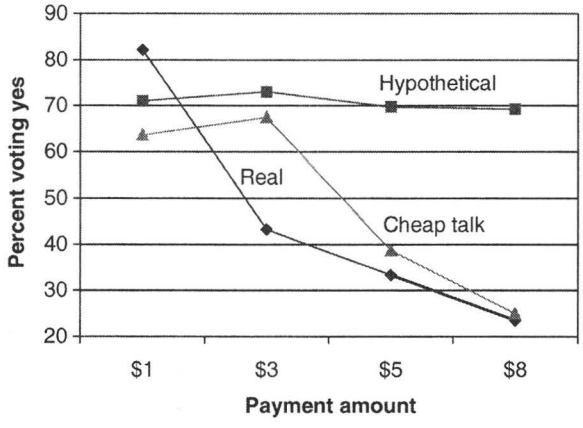


Fig. 1. Referenda on contribution to scholarship.

Table 3
Tests of equal proportions voting yes by payment amount (Fisher's exact p-value, two-sided test)

Null hypothesis	Hypothetical	Cheap talk	Real
$1 = $3	1.000	0.809	0.002
$1 = $5	1.000	0.079	0.000
$1 = $8	1.000	0.001	0.000
$3 = $5	0.796	0.019	0.458
$3 = $8	0.814	0.000	0.229
$5 = $8	1.000	0.221	0.529

amount, with 69–73% of participants voting yes.[4] At the 95% confidence level we cannot reject the hypothesis of no difference in voting among the four hypothetical referenda (Table 3).[5] However, in the real referenda, percent voting yes decreased monotonically as payment amount increased, from 82% at the $1 level to 24% at the $8 level (percent voting yes is significantly greater at the $1 payment level than at the other three payment levels, Table 3). Clearly, participants were sensitive to cost when spending real money. In the cheap talk referenda, percent voting yes decreased with payment amount across the $3, $5, and $8 amounts (with the $3 percentage yes being significantly greater than the $5 and $8 percentages yes, Table 3), but the $1 percentage yes was slightly below and not significantly different from the $3 percentage yes.

Examining the tendency toward hypothetical bias, percentages voting yes in the hypothetical referenda were significantly greater than in the real referenda at the $3, $5, and $8 levels (Table 4).

[4] To test whether a yes vote of about 70% was an invariable result for public goods with our subject population, we also ran a hypothetical referendum with 19 participants on donating $8 to a different scholarship fund, one usable at universities throughout the U.S., and observed a 47% yes vote. This extra referendum suggests sensitivity to the nature of the good in the hypothetical condition. The difference in percentages voting yes between the UMass scholarship fund and this other scholarship fund was marginally significant (the Fisher's exact two-sided test yields a p-value of 0.065).

[5] Two-sided tests of hypotheses are reported herein using Fisher's exact test. One-sided tests, of course, yield lower probability levels, but reach the same conclusions for our data at the 0.05 significance level, except for the comparison of $1 and $5 in the cheap talk treatment, for which the probability level drops from 0.079 with a two-sided test to 0.040 with a one-sided test.

Table 4
Tests of equal proportions voting yes by treatment (Fisher's exact p-value, two-sided test)

Null hypothesis	Payment amount			
	$1	$3	$5	$8
Real = hypothetical	0.370	0.018	0.006	0.001
Real = cheap talk	0.154	0.042	0.791	1.000
Hypothetical = cheap talk	0.601	0.632	0.023	0.000

At the $1 level, the percentage voting yes was actually greater in the real referendum than in the hypothetical referendum, though not significantly so.

The cheap talk script had the desired effect of eliminating hypothetical bias at the higher payment levels, dropping percent voting yes to within one percentage point of what it was in the real referendum at the $8 level, and to within 6 percentage points at the $5 level.[6] At both of these payment amounts, the hypothetical percentages yes are significantly greater than the corresponding cheap talk percentages (Table 4). But at the $3 payment level the cheap talk script had a very modest effect, producing a percentage yes only 6 points below the hypothetical percentage yes, and significantly greater than the real percentage yes (Table 4). Finally, at the $1 level no correction was needed and the correction with the cheap talk script was not statistically significant.

5. Discussion

This study joins a number of prior split-sample studies demonstrating hypothetical bias in elicitations of willingness to pay by comparing responses from hypothetical and real payment conditions. Two of these studies [5,7] used referenda to estimate values of public goods, but employed only a $10 payment amount. We extended their work by testing for bias in several referenda differing in payment amount. Among the four payment amounts ($1, $3, $5 and $8), the percent voting in favor in the hypothetical referenda remained nearly constant, whereas when the referenda were for real the percent in favor steadily dropped as the amount was raised, causing hypothetical bias to increase with payment amount.

The sensitivity of the vote to payment amount in the real referenda was expected, but the insensitivity to payment amount in the hypothetical referenda was not, as previous studies of hypothetical responses have found that proportion responding yes drops as bid level rises. Such sensitivity has been found for a private good at bid levels ($1 and $5) similar to those used here (e.g., [1]). And dichotomous choice contingent valuation studies of public goods regularly show a significant negative relation between stated willingness to pay and bid level [for contingent

[6] The effect of the script was also tested in three additional $8 referenda, which differed from the standard $8 referenda in that the respondents were informed before the vote that they would be asked after the vote to reveal and explain their votes to others in the room. In these referenda, 63%, 45% and 43% voted yes in the hypothetical, cheap talk, and real referenda, respectively. Thus, anticipation that the votes would be discussed tended to narrow the difference between the hypothetical and real referenda, compared with the secret ballots, but the effect of the script was again to correct for the bias. For more detail, see our discussion paper "Obtaining Unbiased Contingent Values: Further Tests of Entreaties to Avoid Hypothetical Bias," available at: http://www.fs.fed.us/rm/value/docs/dp-01-1.pdf.

referenda examples, see [4,14]. However, unlike our study, most contingent valuation studies use a wide range of payment amounts. Thus, the lack of sensitivity to payment amount that we found cannot be generalized to higher payment amounts.

Why was the hypothetical vote invariant to payment amount, thereby causing the substantial hypothetical bias we observed? Four possibilities come to mind, the first three of which are suggested in the cheap talk script. First, some participants in hypothetical referenda may try to estimate their real willingness to pay but may over-estimate. Second, some may ignore their budget constraint, thereby responding as if they faced none. Third, some may use their vote to indicate whether or not they think the cause is worthy of support. Fourth, some may use their vote to indicate whether or not they think such a referendum should be held for real. In essence, participants acting according to the latter three of these possibilities are answering different questions from the one they were asked in the hypothetical referendum.

The Cummings and Taylor cheap talk script worked remarkably well at the two larger payment amounts, lowering the percent voting in favor to within 1 ($8 payment) and 6 ($5 payment) percentage points of what it was in the real referenda. This result is similar to those obtained by Cummings and Taylor in their public good referenda, which used a $10 payment with a similar subject pool. However, the script significantly under-corrected at the $3 payment level, lowering percentage voting in favor only 6 percentage points below that of the hypothetical condition. What could explain these conflicting results?

The cheap talk script tries to bring to the participant's attention factors that are presumably considered in real referenda but that may not readily come to mind in hypothetical referenda. In discussing these factors the script may help participants to focus on the question asked (rather than on ancillary questions that they might prefer to answer), and then helps them to think realistically. There is no evidence that the script did not achieve this at the $8 and $5 payment levels. One explanation for the script's limited effect at the $3 payment level is that some participants may have thought it implausible that they would vote against the referendum at such a low payment. That is, they may not have envisioned significant bias at that level, leaving them less susceptible to the script.

The Cummings and Taylor script is long, and would thus be difficult to use on the phone or in the mail. A much shorter version of the entreaty used by Loomis et al. [13], which did not specifically educate respondents about hypothetical bias, helped reduce hypothetical bias but under-corrected on average. Based simply on the differences between the Loomis et al. and the Cummings and Taylor scripts, it appears that the entreaty has a greater impact if it not only instructs participants to act as if the response were for real, but also reports that responses to a hypothetical payment question are often very different from those to a real payment question and then explains why such a difference might occur.

To summarize, ours and earlier tests of the entreaty method suggest that: (1) entreaties have an effect, although the effect may be modest depending on its wording and on the details of the valuation question; (2) just urging respondents to act as if they were really spending their money helps, but is insufficient; (3) to be effective the script must make hypothetical bias salient to the respondent; (4) the cheap talk script appears to increase in effectiveness as payment level, and correspondingly hypothetical bias, increase—though the evidence of this is limited to the relatively modest payment levels used here and by Cummings and Taylor; and (5) under-correction can happen at low payment amounts.

Future research must determine (1) which parts of the Cummings and Taylor script are essential and which are not; (2) what the effect of the script is at higher referendum payment levels; and (3) whether a different script would be more effective at very low payment levels. Of course, development of cheap talk scripts would be aided by a theory of hypothetical bias, explaining why it occurs. Development of such a theory may require moving from studies of aggregate behavior to research at the individual level. Work at the individual level would determine whether hypothetical bias occurs in all respondents, or only in some. It would then determine whether entreaties affect only respondents subject to bias (causing accurate correction for each such respondent), or all respondents (causing a set of under-corrections balanced by a set of over-corrections).

References

[1] K. Blumenschein, M. Johannesson, G.C. Blomquist, B. Liljas, R.M. O'Connor, Experimental results on expressed certainty and hypothetical bias in contingent valuation, Southern Econom. J. 65 (1998) 169–177.
[2] R.R. Boyce, T.C. Brown, G.D. McClelland, G.L. Peterson, W.D. Schulze, Experimental evidence of existence value in payment and compensation contexts, in: K.J. Boyle, T. Heekin (Eds.), Western Regional Research Project W-133: Benefits and Costs in Natural Resources Planning, University of Maine, Orono, ME, 1989.
[3] T.C. Brown, P.A. Champ, R.C. Bishop, D.W. McCollum, Which response format reveals the truth about donations to a public good?, Land Econom. 72 (1996) 152–166.
[4] R.T. Carson, W.M. Hanemann, R.J. Kopp, J.A. Krosnick, R.C. Mitchell, S. Presser, P.A. Ruud, V.K. Smith, M. Conaway, K. Martin, Referendum design and contingent valuation: the NOAA panel's no-vote recommendation, Rev. Econom. Statist. 80 (1998) 484–487.
[5] R.G. Cummings, S. Elliott, G.W. Harrison, J. Murphy, Are hypothetical referenda incentive compatible?, J. Polit. Econom. 105 (1997) 609–621.
[6] R.G. Cummings, G.W. Harrison, E.E. Rutstrom, Homegrown values and hypothetical surveys: is the dichotomous choice approach incentive-compatible?, Amer. Econom. Rev. 85 (1995) 260–266.
[7] R.G. Cummings, L.O. Taylor, Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method, Amer. Econom. Rev. 89 (1999) 649–665.
[8] M. Dickie, A. Fisher, S. Gerking, Market transactions and hypothetical demand: a comparative study, J. Amer. Statist. Assoc. 82 (1987) 69–75.
[9] C.L. Eavey, G.J. Miller, Bureaucratic agenda control: imposition or bargaining, Amer. Polit. Sci. Rev. 78 (1984) 719–733.
[10] J.P. Hoehn, A. Randall, A satisfactory benefit cost indicator from contingent valuation, J. Environ. Econom. Management 14 (1987) 226–247.
[11] M.J. Kealy, J.F. Dovidio, M.L. Rockel, Accuracy in valuation is a matter of degree, Land Econom. 64 (1988) 158–171.
[12] J.A. List, Do explicit warnings eliminate the hypothetical bias in elicitation procedures? Evidence from field auctions for sportscards, Amer. Econom. Rev. 91 (2001) 1498–1507.
[13] J. Loomis, T. Brown, B. Lucero, G. Peterson, Improving validity experiments of contingent valuation methods: results of efforts to reduce the disparity of hypothetical and actual WTP, Land Econom. 72 (1996) 450–461.
[14] J. Loomis, A. Gonzalez-Caban, R. Gregory, Do reminders of substitutes and budget constraints influence contingent valuation estimates?, Land Econom. 70 (1994) 499–506.
[15] H. Neill, R. Cummings, P. Ganderton, G. Harrison, T. McGuckin, Hypothetical surveys and real economic commitments, Land Econom. 70 (1994) 145–154.
[16] T. Romer, H. Rosenthal, Bureaucrats versus voters: on the political economy of resource allocation by direct democracy, Quart. J. Econom. 93 (1979) 563–587.

Explaining the Discrepancy Between Intentions and Actions: The Case of Hypothetical Bias in Contingent Valuation

Icek Ajzen
University of Massachusetts, Amherst

Thomas C. Brown
U.S. Forest Service, Rocky Mountain Research Station, Fort Collins, Colorado

Franklin Carvajal
University of Massachusetts, Amherst

An experiment was designed to account for intention-behavior discrepancies by applying the theory of planned behavior to contingent valuation. College students (N = 160) voted in hypothetical and real payment referenda to contribute $8 to a scholarship fund. Overestimates of willingness to pay in the hypothetical referendum could not be attributed to moderately favorable latent dispositions. Instead, this hypothetical bias was explained by activation of more favorable beliefs and attitudes in the context of a hypothetical rather than a real referendum. A corrective entreaty was found to eliminate this bias by bringing beliefs, attitudes, and intentions in line with those in the real payment situation. As a result, the theory of planned behavior produced more accurate prediction of real payment when participants were exposed to the corrective entreaty.

Keywords: *attitude; contingent valuation; hypothetical bias; intention; theory of planned behavior*

It is a common observation that people often fail to act in accordance with their stated intentions. The study reported in this article examined possible reasons for the discrepancy between expressed willingness to perform a behavior and its actual performance. Empirical research on this problem can be traced to LaPiere's (1934) classic study on racial prejudice. A Chinese couple stopped at more than 250 restaurants, coffee shops, hotels, motels, and inns while touring the United States and was admitted and received service without hesitation in 95% of the instances; yet, in response to a letter of inquiry, 92% of the establishments replied that they would not accept members of the Chinese race. Subsequent investigations have produced similar findings. For example, among female college students who indicated intentions to release their photos with an African American male for a variety of purposes, approximately 40% failed to follow through (Linn, 1965), and in the health domain, between 26% and 57% of respondents failed to carry out their intentions to use condoms, to undergo a cancer screening, or to exercise (Sheeran, 2002).

In the present study, we investigated the intention-behavior discrepancy in the context of contingent value measurement. Contingent valuation is a popular tool for assessing the monetary value of goods not traded in the market place (for reviews, see Cummings, Brookshire, & Schulze, 1986; Mitchell & Carson, 1989). Respondents in a survey are asked to indicate their willingness to pay for a certain good in a hypothetical or contingent market. The monetary value of the good in question is measured by aggregating these willingness-to-pay judgments in the relevant population. Unfortunately, scores of contingent valuation surveys conducted in recent years have revealed that many factors bias the amount of money participants indicate they would be willing to pay, thus jeopardizing the method's validity (see Hoehn & Swanson, 1988; Mitchell & Carson, 1989, for reviews).

Authors' Note: Franklin Carvajal is now at the Virginia Polytechnic Institute and State University in Blacksburg, Virginia. Correspondence concerning this article should be addressed to Icek Ajzen, Department of Psychology, Tobin Hall, Box 37710, University of Massachusetts, Amherst, MA 01003-7710; e-mail: aizen@psych.umass.edu.

PSPB, Vol. 30 No. 9, September 2004 1108-1121
DOI: 10.1177/0146167204264079
© 2004 by the Society for Personality and Social Psychology, Inc.

Prominent among recent concerns is mounting evidence for the so-called hypothetical bias, an overestimation of willingness to pay in hypothetical or contingent markets compared to actual payment in otherwise identical real cash markets (see Blumenschein, Johannesson, Blomquist, Liljas, & O'Connor, 1998; Brown, Champ, Bishop, & McCollum, 1996; Cummings, Harrison, & Rutstrom, 1995; Loomis, Brown, Lucero, & Peterson, 1996; Neill, 1995). For example, in different conditions of a recent experiment (Brown, Ajzen, & Hrubes, 2003), college students voted in a referendum to contribute $1, $3, $5, or $8 to a scholarship fund either in a hypothetical or a real payment context. Except for the $1 condition, in which the proportion of real yes votes exceeded the proportion of hypothetical yes votes by 10%, all other payment amounts revealed the usual hypothetical bias; the excess of hypothetical as opposed to real yes votes ranged from 30% in the $3 condition to 48% in the $8 condition.

EXPLAINING INTENTION-BEHAVIOR DISCREPANCIES

Expressions of willingness to pay in a hypothetical situation can be likened to behavioral intentions (Ajzen & Driver, 1992)[1] and hypothetical bias to a discrepancy between intention and behavior. The study reported in this article explored two possible reasons for hypothetical bias. One explanation was offered by Campbell (1963), who maintained that hypothetical and actual responses are both indicators of the same underlying latent disposition. People with highly positive dispositions would be expected to respond favorably in hypothetical as well as real contexts, whereas people with highly negative dispositions would be expected to respond negatively in both contexts. The discrepancy between intentions and behavior can, according to Campbell, be traced to individuals with moderate dispositions who respond favorably in the hypothetical context but unfavorably in the more demanding real context. The hypothetical bias in this view is more apparent than real, or—in Campbell's terms—it is a case of pseudo-inconsistency.

The second explanation attributes the discrepancy between intentions and behavior to substantive differences between hypothetical and real contexts. Early accounts traced these differences to the way in which people construe the situation, to the possibility that symbolic representations may be very different from representations in real-life situations (Blumer, 1955). In fact, LaPiere's (1934) study of racial prejudice was motivated by the desire to show that "symbolic attitudes are seldom more than a verbal response to a symbolic situation" (p. 230) and that "there is no necessary correlation between speech and action, between response to words and to the realities they symbolize" (p. 231).

In a related but more contemporary interpretation, it may be argued that salient features of a behavioral situation often activate beliefs about a behavior that are different from the beliefs that are activated in the hypothetical situation in which verbal questionnaire responses are elicited (Ajzen & Sexton, 1999). To explain the intention-behavior discrepancy shown in hypothetical bias, one would have to assume that the symbolic situation activates more favorable or fewer unfavorable considerations than does the behavioral situation. Behavior consistent with intentions would only be expected when beliefs in the two situations are congruent, that is, when they are the same or at least equally favorable or unfavorable.

The latent disposition hypothesis advanced by Campbell (1963) and the alternative belief disparity hypothesis are not mutually exclusive, and both processes could be at work. According to the first, hypothetical bias is due to a moderately favorable latent disposition among many respondents, a disposition that produces willingness to pay in the symbolic context but not in the manifest context. This latent disposition hypothesis implies that individuals who display hypothetical bias should have less favorable beliefs and attitudes regarding payment for the public good under consideration than individuals who are willing to pay in both contexts but more favorable beliefs and attitudes than individuals who are willing to pay in neither. Beliefs and attitudes regarding payment for a certain good represent the stable underlying disposition individuals bring to the situation; these beliefs and attitudes are the same in symbolic and behavioral contexts.

The belief-disparity hypothesis suggests that beliefs and attitudes are not invariant across context. Although they contain a stable core, beliefs and attitudes are assumed to be strongly influenced by salient contextual cues. The implication of this account is that we would expect more favorable beliefs and attitudes in the symbolic than in the behavioral context.

PREDICTING WILLINGNESS TO PAY: THE THEORY OF PLANNED BEHAVIOR

To study the competing explanations of hypothetical bias, it is necessary to assess behavior-relevant beliefs and attitudes prior to observation of willingness to pay under hypothetical and real payment conditions. Assessment of beliefs and attitudes in the present study was guided by the theory of planned behavior (Ajzen, 1988, 1991), which has been applied successfully to a diverse range of behavioral domains (see Ajzen, 2001b; Armitage & Conner, 2001). Briefly, according to the theory, human action is influenced by three major factors: a favorable or unfavorable evaluation of the behavior (attitude toward the behavior), perceived social pressure to perform or

not perform the behavior (subjective norm), and self-efficacy in relation to the behavior (perceived behavioral control). In combination, attitude toward the behavior, subjective norm, and perception of behavioral control lead to the formation of a behavioral intention. As a general rule, the more favorable the attitude and subjective norm, and the greater the perceived behavioral control, the stronger should be the person's intention to perform the behavior in question. Finally, given a sufficient degree of actual control over the behavior, people are expected to carry out their intentions when the opportunity arises (see also Gollwitzer, 1993; Triandis, 1977). Intention is thus assumed to be an immediate antecedent of behavior. However, because many behaviors pose difficulties of execution that can limit volitional control, it is useful to consider perceived behavioral control in addition to intention. To the extent that people are realistic in their judgments of a behavior's difficulty, a measure of perceived behavioral control can serve as a proxy for actual control and contribute to the prediction of the behavior in question (see Ajzen, 1991). Perceived behavioral control can thus influence behavior indirectly via intentions, and as a proxy for actual control, it can have a direct link to behavior. A schematic representation of the theory is shown in Figure 1 below.

When applied to contingent valuation in the context of a referendum, the behavior of interest is the vote (yes or no) to contribute money to a worthy goal. The three major determinants of this behavior—attitudes toward voting yes, subjective norms, and perceptions of behavioral control—are traced to corresponding sets of behavior-relevant beliefs. Consistent with an expectancy-value model (Feather, 1982; Fishbein, 1963), attitude toward voting yes is assumed to be determined by beliefs about the consequences of doing so, each belief weighted by the subjective value of the consequence in question (Fishbein, 1963, 1967; Fishbein & Ajzen, 1975). A similar logic applies to the relation between normative beliefs and subjective norm and the relation between control beliefs and perceived behavioral control. Normative beliefs refer to the perceived behavioral expectations of such important referent individuals or groups as the person's family, friends, teachers, and coworkers. These normative beliefs—in combination with the person's motivation to comply with the different referents—determine the prevailing subjective norm regarding the vote. Finally, control beliefs have to do with the perceived presence of factors that can facilitate or impede performance of a behavior. It is assumed that the perceived power of each control factor to impede or facilitate voting yes contributes to perceived control over this behavior in direct proportion to the person's subjective probability that the control factor is present.

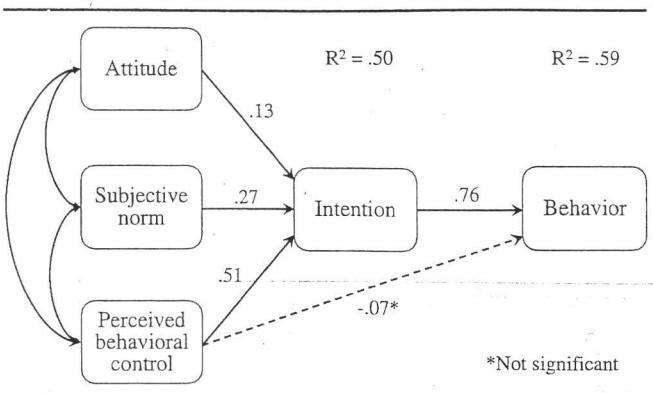


Figure 1 Prediction of voting in real referendum: Path analysis for the theory of planned behavior—total sample.

The present study, then, was designed to investigate beliefs, attitudes, and intentions of individuals confronted with the decision to vote yes or no in a referendum under hypothetical and real payment conditions. Measuring these variables should permit not only prediction of actual votes but also should enable us to gain a better understanding of factors responsible for the hypothetical bias that is observed in contingent valuation surveys, and for intention-behavior discrepancies in general.

REDUCING THE INTENTION-BEHAVIOR GAP

Investigators have tried to adjust for hypothetical bias in a number of ways: by including the use of real payment bids for comparable goods as a way to calibrate the hypothetical bids (Blackburn, Harrison, & Rutstrom, 1994; Fox, Shogren, Hayes, & Kliebenstein, 1998), by asking respondents how certain they are that they really would be willing to pay the indicated amount and using the certainty response to adjust the bids (Champ, Bishop, Brown, & McCollum, 1997; Li & Mattsson, 1995), and by asking respondents to consider their budgetary constraints before providing their hypothetical bids (Loomis et al., 1996; Neill, Cummings, Ganderton, Harrison, & McGuckin, 1994; Posavac, 2001).

In the present study, we used a direct attack on hypothetical bias, entreating participants in a contingent valuation survey to respond as if they were really spending their money. A thorough script designed to remove bias from a hypothetical referendum by means of such a corrective entreaty has been developed and tested with notable success by Cummings and Taylor (1999). In the eight-paragraph script, participants are not only encouraged to act as if the referendum were for real but also are explicitly told about the problem of hypothetical bias and are given possible reasons for the phenomenon. The study by Brown et al. (2003) mentioned earlier replicated and extended this experiment with similarly

impressive results. In this study, college students voted in a referendum to contribute money to a scholarship fund either with or without the entreaty to avoid hypothetical bias and, at the higher payment amounts, the contingent valuation estimates under hypothetical payment conditions were very close to those obtained for real payment.

The belief and attitude data collected in the present study permitted us to explore possible reasons for the effectiveness of the corrective entreaty. Specifically, we expected that, following the entreaty, beliefs and attitudes under hypothetical payment conditions would be comparable to the beliefs and attitudes of respondents in the real payment condition (without any entreaty). As a result, these beliefs and attitudes should become more accurate predictors of actual behavior.

HYPOTHESES

Willingness-to-pay responses were obtained in the context of a referendum, arguably the most appropriate contingent valuation format for valuing public goods (see Hoehn & Randall, 1987). Unlike most previous experiments, in which participants voted either in a hypothetical payment referendum or in a real payment referendum, most of our participants were asked to vote first under hypothetical payment conditions and then again under real payment conditions. This enabled us to identify individuals who displayed the hypothetical bias and to compare them to individuals who did not. The main purpose of the study was to explore the reasons for hypothetical bias, with a secondary aim of examining the effect of a corrective entreaty on this bias. The considerations and predictions discussed above are summarized in the following hypotheses:

Hypothetical bias: As in previous studies, we expect that—without a corrective entreaty—individuals are more likely to vote in favor of a payment referendum under hypothetical than under real payment conditions. This difference is expected to be eliminated or at least greatly reduced when the corrective entreaty is used.

Latent disposition hypothesis: The critical comparison for this hypothesis involves respondents who (without a corrective entreaty) vote yes in the hypothetical referendum and then vote no in the real referendum (i.e., display hypothetical bias) versus respondents who vote yes in both referenda. The latent disposition hypothesis predicts that the latter respondents will have more favorable beliefs, attitudes, and intentions with respect to voting yes than the former.

Belief disparity hypothesis: The critical comparison for this hypothesis is between participants whose beliefs, attitudes, and intentions are assessed under hypothetical payment conditions as opposed to participants whose beliefs, attitudes, and intentions are assessed under real payment conditions. In the absence of a corrective entreaty, the belief disparity hypothesis predicts that beliefs, attitudes, and intentions with respect to voting yes will be more favorable when participants anticipate a hypothetical referendum than when they anticipate a real referendum. By way of contrast, when beliefs, attitudes, and intentions are assessed after exposure to the corrective entreaty, there should be few differences between the two types of referenda because the entreaty should influence beliefs, attitudes, and intentions in the hypothetical situation such as to make them comparable to the beliefs, attitudes, and intentions in the real payment situation.[2]

Prediction of intended and actual votes: In line with the theory of planned behavior, it is expected that intentions to vote yes in a referendum can be predicted from attitudes, subjective norms, and perceptions of behavioral control with respect to voting yes. Furthermore, exposure to a corrective entreaty is expected to affect intentions through its influence on the three antecedent variables. As a result, prediction of intentions should be accurate irrespective of the presence or absence of an entreaty. In contrast, because intentions to vote yes are conceptually analogous to hypothetical votes (Ajzen & Driver, 1992), hypothetical bias is expected to produce a discrepancy between intentions and actual voting in the real referendum. The theory of planned behavior should therefore provide a more accurate prediction of real votes following a corrective entreaty than without it.

METHOD

The public good in the present study involved a proposition to donate $8 to a college scholarship fund. The proposition was identical to one used in a previous investigation (Brown et al., 2003). The participants—University of Massachusetts (UMass) undergraduates—were told of the need to provide financial support to students from poor families who cannot afford to pay the university's tuition and fees. It was explained that financial aid provided by the state is insufficient and that the university maintains a general scholarship fund financed by private donations. Participants in small groups were told that if everyone present were to contribute $8, these contributions would be added to the University Scholarship Fund and would be used to help deserving students to attend the university. Specifically, the proposition took the following form:

PROPOSITION

> Everyone here in this room will contribute $8 to the University of Massachusetts–Amherst Scholarship Fund. The contribution is to be used to provide scholarship aid to needy and deserving students at the University of Massachusetts–Amherst.

It was explained that if more than 50% vote yes on this proposition then everybody will pay $8 and the money will be collected and given to the university for credit to the scholarship fund, but if 50% or fewer of the group members vote yes, then no one will pay $8. The real and hypothetical versions of the referendum differed only in the tense used to describe the outcomes. For example, the first sentence of the real version read, "If more than 50% of you vote YES on this proposition, all of you pay $8.— I will collect $8 from each of you . . . ," whereas the first sentence of the hypothetical version read, "If more than 50% of you were to vote YES on this proposition, all of you would pay $8—I would collect $8 from each of you . . . ."

Procedure

One hundred sixty UMass undergraduates (76% female) participated in the main study in groups of 6 to 10 members. They were paid a $10 compensation and they also received experimental course credit for their participation. After all group members had arrived, they were informed that there were two unrelated parts to the experiment, one involving questionnaire tasks, the other a referendum vote. In the first part, which lasted 45 to 55 min, participants first engaged in an assignment unrelated to the present study (rating fictitious job applicants on a variety of attributes and completing a number of personality scales). In addition, several demographic characteristics were assessed: gender, age, ethnicity, socioeconomic status, political affiliation, and religious affiliation. Because none of the demographic characteristics had a significant effect on the results, they will not be further considered.

When all group members had completed the first part, they were paid $10 for the work they had done and were told that they would now participate in the referendum vote. Instructions for the referendum were read to the group and the text of the referendum also was projected onto a screen. In some conditions of the experiment, the participants were then given paper ballots on which they recorded their votes (yes or no) in secret. After handing in their ballots, they were asked to complete a questionnaire designed to assess the variables in the theory of planned behavior. In other conditions, the participants first completed the theory of planned behavior questionnaire and then proceeded to vote on the proposition. In none of the real payment referenda did a majority of group members vote in favor of making the donation. It was therefore not necessary to collect money from the participants. At the conclusion, all participants were given an explanation of the study, were thanked for their participation, and were offered the address for making donations to the University Scholarship Fund if they desired to do so.

Experimental Conditions

For the referendum part of the study, groups of participants were assigned at random into one of five experimental conditions. In each condition, the referendum was explained first, but the remaining aspects varied, as follows:

1. Q-R. In the first condition, participants completed the theory of planned behavior questionnaire (Q) after receiving information about the referendum, informing them that the referendum was for real. This was followed by a vote under real payment conditions (R), that is, participants were told that they would have to pay $8 if the majority voted in favor of the proposition.

2. Q-H-R. In the second condition, completion of the theory of planned behavior questionnaire (Q) was followed first by a referendum under hypothetical payment conditions (H). Participants were told that they would not actually have to pay, even if a majority voted in favor of the proposition, but they were asked to respond as though the vote involved real cash payments. Once they had handed in their ballots, they were told that they would now participate in a real referendum where they will actually have to pay $8 if the majority votes yes and then proceeded to the second referendum under real payment conditions (R).

3. H-R-Q. To control for the possible effect of administering the theory of planned behavior questionnaire prior to voting on a referendum, participants in the third condition completed the questionnaire at the end. In this condition, the description of the referendum was followed immediately by a hypothetical vote (H), then a vote under real payment conditions (R), and finally administration of the questionnaire (Q). The questionnaire was identical to the questionnaire given in the first two conditions except that most items were worded in the past tense.

4. E-Q-H-R. The fourth condition was identical to the second except that the description of the referendum contained the corrective entreaty (E) designed to eliminate hypothetical bias (described below).

5. E-H-R-Q. In the final condition, the theory of planned behavior questionnaire was again moved to the end to control for the possible impact of completing the questionnaire prior to casting a ballot.

Corrective Entreaty

In conditions containing the corrective entreaty, participants were told of the problem of hypothetical referenda—that when money payments are at issue, people tend to vote differently under hypothetical than under real payment conditions. Results of past studies were presented showing a higher percentage voting yes in a hypo-

thetical referendum than in a real referendum. The reasons for this difference were then discussed, including the notion that in a referendum for a worthy cause, people tend to focus on doing good and tend to ignore the cost to them when it is hypothetical, but they temper that inclination with considerations of other options for their money when the referendum is for real. The entreaty ended with the following request: "I ask you to vote just exactly as you would vote if you were really going to face the consequences of your vote: which is to pay money if the proposition passes."[3]

Theory of Planned Behavior Questionnaire

Based on a pilot study with 46 UMass undergraduates, a seven-page questionnaire was developed to measure the constructs in the theory of planned behavior (see Ajzen, 2001a, for a description of questionnaire development). Participants were first given instructions on use of the questionnaire's 7-point bipolar adjective scales. There was no systematic order to the items designed to assess the theory's different constructs.

Direct attitude. The stem, "For me to contribute $8 to the UMass Scholarship Fund would be" was rated on a series of ten 7-point evaluative semantic differential scales with the following endpoints: *harmful-beneficial, undesirable-desirable, unpleasant-pleasant, good-bad, reasonable-unreasonable, comfortable-uncomfortable, negative-positive, fair-unfair, enjoyable-unenjoyable, wrong-right.* High scores were assigned to the positive end of each scale and the mean response across the 10 scales was used as a direct measure of attitude toward voting yes on the referendum. The internal reliability of this measure, as assessed by Cronbach's alpha, was .90.

Belief-based attitude. Two questions were asked with respect to each of five outcomes of voting to contribute $8 to the UMass Scholarship Fund identified in the pilot study. The five outcomes dealt with making more money available for scholarships to needy students, forcing others in the group to make a donation, making a difference for the better in someone's life, donating money that could be used for other things, and enabling more needy students to attend UMass. First, participants were asked to evaluate each outcome on a 7-point *extremely good-extremely bad* scale. Second, to assess belief strength, they were asked to rate the likelihood that voting to contribute $8 to the UMass Scholarship Fund would produce each of the five outcomes on a 7-point *extremely unlikely-extremely likely* scale. Based on an optimal scaling analysis that determines the proper scoring scheme (Ajzen, 1991), belief strength and outcome evaluation were both scored in a bipolar fashion, from –3 (*unlikely, bad*) to +3 (*likely, good*). To produce a belief-based estimate of attitude, belief strength and outcome evaluation, measures were multiplied and the resulting products were summed, in accordance with the expectancy-value model discussed in the Introduction.[4] The correlation between the direct and this belief-based measure of attitude was .50 ($p < .001$).

Direct subjective norm. Three items, interspersed among other questions, assessed subjective norms. Respondents rated, on 7-point scales, the statements that most people who are important to them would want them to vote to contribute $8 to the UMass Scholarship Fund (*strongly agree-strongly disagree*) and would think that they should vote to contribute (*should-should not*). The third item asked them to rate the statement that most people who are close to them would themselves vote to contribute $8 to the UMass Scholarship Fund (*definitely true-definitely false*). The mean score across the three items, with a Cronbach's alpha coefficient of .85, served as a direct measure of subjective norm.

Belief-based subjective norm. The pilot study had identified four normative referents of relevance: parents and family, friends and fellow students, other participants voting on the referendum, and UMass faculty and administrators. With respect to each of these four referents, two items assessed normative belief strength and motivation to comply. For example, the statement, "My friends and fellow students would think that I should contribute $8 to the UMass Scholarship Fund" was rated on a 7-point *extremely unlikely-extremely likely* scale to produce a measure of normative beliefs strength, and to assess motivation to comply, participants rated, on a 7-point *strongly disagree-strongly agree* scale, the statement, "Generally speaking, I want to do what my friends and fellow students think I should do." Optimal scaling analysis resulted in bipolar scoring (–3 to +3) for normative belief strength but unipolar (1 to 7) scoring for motivation to comply. The products of normative belief strength and motivation to comply were summed across the four referents to obtain a belief-based estimate of subjective norm. This measure had a correlation of .67 ($p < .001$) with the direct subjective norm measure.

Direct perceived behavioral control. Two items, appearing at different points in the questionnaire, were designed to assess perceived behavioral control. Respondents rated how possible they thought it was for them to vote to contribute $8 to the UMass Scholarship Fund (*possible-impossible*) and that the decision to contribute was completely up to them (*strongly agree-strongly disagree*). Perceived control, as measured by the second item, was found to be quite high, with relatively little variance among respondents; that is, most participants believed that the decision to contribute $8 to the UMass Scholarship fund was up to them. On the 7-point scale used, the mean response was 6.20 ($SD = 1.33$). In fact, this item

revealed virtually no correlation with any of the other variables assessed, and it did not correlate with the first perceived behavioral control item ($r = -.01$). Consequently, only responses to the first item were used as a direct measure of perceived behavioral control.

Belief-based perceived behavioral control. Four control factors were identified in the pilot study as potentially facilitating or hindering a positive vote in the referendum: whether the person could afford to make the contribution, could be sure that the contribution would indeed go to needy students, would not have to draw on regularly available funds, and had no other immediate need for the money. To assess control-belief strength, participants rated each control factor on a 7-point scale ranging from *definitely false* to *definitely true.* For example, "I can afford to contribute $8 to the UMass Scholarship Fund" was rated on this scale. A second item assessed the perceived power of the control factor. Thus, respondents rated, on a 7-point *strongly disagree-strongly agree* scale, the statement, "Being able to afford to contribute $8 to the UMass Scholarship Fund would enable me to vote in favor of the referendum." Optimal scaling analysis led to unipolar scoring (1 to 7) for control beliefs as well as perceived power of control factors. Control-belief strength and perceived power were multiplied for each control factor and the summed product terms constituted the belief-based measure of perceived behavioral control. The correlation between direct and belief-based measures of perceived control was .56 ($p < .001$).

Intention. Two 7-point items, separated by other measures, were designed to assess intentions to vote to contribute $8 to the UMass Scholarship Fund. One item asked respondents whether they intended to do so (*not at all-definitely*), the other whether they planned to do so (*definitely true-definitely false*). Cronbach's alpha for these two items was .93. Responses to the two items were averaged to obtain a measure of intention.[5] In addition, a binary intention measure was obtained by asking respondents, at the end of the questionnaire, whether they will vote yes or no in the referendum.

RESULTS

Hypothetical Bias and Corrective Entreaty

Our first order of business was to confirm the presence of hypothetical bias in the contingent valuation and to examine the ameliorating effect of the corrective entreaty. Table 1 shows the number of participants in each of the experimental conditions and the proportion who voted in favor of the proposition in the hypothetical and real payment referenda. Because neither hypothetical nor real voting behavior were affected by order of questionnaire administration, order was collapsed in these analyses. We first compared hypothetical to real votes in the three conditions that did not include a corrective entreaty, that is, Conditions 1, 2, and 3. As expected, the proportion of participants voting yes in the hypothetical referendum (mean of Conditions 2 and 3 = 0.69) was significantly higher than the proportion voting yes in the real referendum (mean of Conditions 1, 2, and 3 = 0.41; Wilkinson matched pairs signed ranks test, $z = 4.15$, $p < .001$). However, the corrective entreaty (mean of Conditions 4 and 5) significantly reduced the proportion of hypothetical yes votes from .69 to .33, $\Pi^2(1, N = 129) = 17.00$, $p < .001$, and following the entreaty, the mean difference between hypothetical and real votes (.33 vs. .28) was no longer significant (Wilkinson's $z = 1.34$, $p = .18$). These findings, then, confirm our first hypothesis: Without a corrective entreaty, participants displayed the usual hypothetical bias, but this bias was largely eliminated by the entreaty.

TABLE 1: Number of Participants and Proportion of Yes Votes in Hypothetical and Real Payment Referenda

	Referendum			
	Hypothetical		*Real*	
Experimental Condition	N	*% Yes*	N	*% Yes*
1. Q-R	—	—	31	.42
2. Q-H-R	34	.71	34	.38*
3. H-R-Q	34	.68	34	.44*
4. E-Q-H-R	32	.41	32	.34
5. E-H-R-Q	29	.15	29	.24

NOTE: Q = theory of planned behavior questionnaire, H = hypothetical referendum, R = real referendum, and E = corrective entreaty. The difference between hypothetical and real referenda is significant.
*$p < .05$.

Prediction of Intentions and Behavior

Intentions, hypothetical votes, and real votes. As expected, hypothetical votes were found to be quite similar to intentions. In fact, hypothetical votes correlated more strongly with intentions than with real votes. This can be seen most clearly by focusing on the binary intention measure among the 129 participants who voted in the hypothetical as well as in the real referendum (Conditions 2 through 5). In the hypothetical referendum, 128 participants (more than 99%) voted in accordance with their intentions (67 yes and 61 no), resulting in a phi correlation coefficient of .99 ($p < .001$). Only one participant expressed an intention to vote yes on the questionnaire and then voted no in the hypothetical referendum.

Correspondence between hypothetical and real votes was much lower, with 103 out of 129 (80%) participants voting the same way in the hypothetical and real referenda (phi = .64, $p < .001$). Consistent with the hypothetical bias documented earlier, the discrepant votes were

asymmetrical. Of 67 participants who voted yes in the hypothetical referendum, 24 (36%) voted no in the real referendum but of 62 who voted no in the former, only 2 (3%) voted yes in the latter.

The discrepancies between intentions and voting in the real referendum were similar to the discrepancies between hypothetical and real votes. Of the 129 participants considered, 104 (81%) voted in accordance with their intentions (phi = .66, $p < .001$). When real votes were inconsistent with intentions, virtually all showed overestimates in terms of intentions such that 24 of 68 respondents (35%) actually voted no when they intended to vote yes, but only 1 of 61 (2%) voted yes after expressing an intention to vote no.

Theory of planned behavior and voting in the real referendum. Path analyses were conducted to test the ability of the theory of planned behavior to predict intentions to vote yes and actual voting in the real referendum. Using EQS version 5.7b (Bentler, 1995), direct paths were specified from attitude, subjective norm, and perceived behavioral control to intention and from intention and perceived behavioral control to behavior.[6] Because one of the two items designed as a direct measure of perceived behavioral control had poor psychometric properties, the belief-based measures of attitude, subjective norm, and perceived behavioral control were used in the path analyses but very similar results were obtained when the direct measures were substituted for the belief-based measures. Also, because the behavioral criterion (yes or no vote) was a dichotomous measure, the robust method for categorical variables was applied. In addition to computing the chi-square for the difference between predicted and obtained values, the fit between structural model and data was evaluated by means of the Bentler-Bonnet normed fit index (BBNFI) and the root mean square error of approximation (RMSEA). The BBNFI estimates the amount of variance explained by the model. It can vary from 0 to 1 and a good fit is indicated by values greater than .95. The RMSEA index compensates for sample size, with low values indicating good fit. Usually, a RMSEA value of .05 or less is considered acceptable.

A path analysis was first performed for the total sample. Table 2 displays the means and standard deviations of attitudes, subjective norms, perceptions of behavioral control, intentions, and actual voting behavior, as well as the correlations among these variables. The fit for the proposed model was found to be excellent. The discrepancy between predicted and obtained covariance structures was not significant, $\chi^2 = 4.14$, $df = 2$, $p = .13$, and the goodness-of-fit indices reached highly satisfactory levels (BBNFI = .99, RMSEA = .04). In Figure 1, it can be seen that all three components—attitude toward voting yes, subjective norm, and perceived behavioral control—

TABLE 2: Means, Standard Deviations, and Correlations for Theory of Planned Behavior Variables: Total Sample ($N = 160$)

	M	SD	*A*	*SN*	*PBC*	*I*
Attitude (A)	10.06	13.11	—			
Subjective norm (SN)	10.09	20.29	.43	—		
Perceived behavioral control (PBC)	95.40	44.19	.53	.53	—	
Intention (I)	4.24	2.26	.51	.59	.72	—
Real vote	0.36	0.48	.39	.35	.41	.61

NOTE: All correlations are significant at $p < .001$.

made significant contributions to the prediction of intentions, and intentions, in turn, had a strong effect on actual voting behavior in the real referendum. Perceived behavioral control did not have a significant direct path to behavior but appears to have influenced voting indirectly by its effect on intention. The model accounted for 50% of the variance in intentions and for 59% of the variance in actual vote.

Following the analysis for the total sample, separate analyses were performed for participants in Condition 2 who did not receive the corrective entreaty (and who completed the questionnaire prior to voting) and for participants in the otherwise identical Condition 4 who were entreated to be realistic in their hypothetical votes.[7] The model was found to fit the data quite well in both conditions of the experiment, although—as might be expected—the fit was somewhat better for participants with than without the corrective entreaty. For participants without the entreaty, the discrepancy between predicted and obtained covariance structures was not significant, $\chi^2 = 3.56$, $df = 2$, $p = .15$, and the BBNFI goodness-of-fit index was at an acceptable level (.96). The RMSEA index, however, was .15, which is above the .05 level considered satisfactory. The same analysis conducted for participants exposed to the corrective entreaty (Condition 4) resulted in a good fit between model and data, $\chi^2 = 1.23$, $df = 2$, $p = .54$, *ns*, and both fit indices were highly satisfactory (BBNFI = .99, RMSEA < .001).

The path coefficients and proportions of explained variance in intentions and behavior are shown in Figure 2. The values to the left of each slash refer to the experimental condition that did not include a corrective entreaty (Condition 2) and the values to the right of the slash to the condition that included the entreaty (Conditions 4). Inspection of Figure 2 reveals that, as expected, the theory afforded good prediction of intentions irrespective of entreaty condition. The difference between the amount of explained variance without entreaty ($R^2 = .64$) and with entreaty ($R^2 = .57$) was not significant ($z = 0.49$, $p = .62$).

The most notable difference between entreaty conditions is related to the proportion of explained behav-

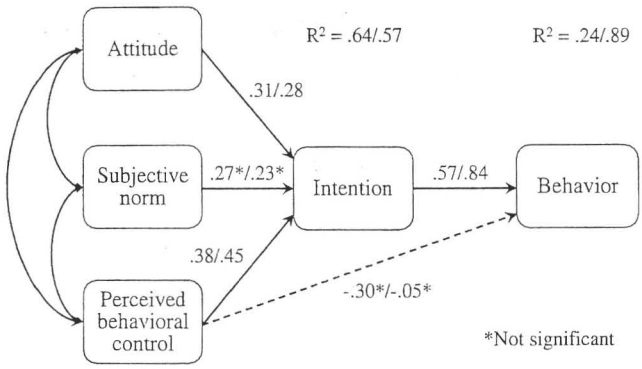


Figure 2 Prediction of voting in real referendum: Path analysis for the theory of planned behavior without corrective entreaty (values before slash) and with corrective entreaty (values after slash).

TABLE 3: Univariate Comparison of Means for Theory of Planned Behavior Variables: Differences Between Real No and Real Yes Votes for Participants Who Voted Yes in the Hypothetical Referendum

	Real Vote = No		*Real Vote = Yes*		
	M	SD	M	SD	F(df = *1, 44*)
Attitude					
Direct	5.99	1.94	5.76	1.18	0.25
Belief-based	13.11	6.90	16.22	11.69	1.08
Subjective norm					
Direct	5.91	0.75	5.98	1.10	0.05
Belief-based	23.79	19.39	25.44	22.35	0.07
Perceived behavioral control					
Direct	6.42	0.77	6.33	1.44	0.06
Belief-based	118.81	47.22	121.96	42.11	0.36
Intention	6.26	0.89	6.15	1.59	0.82

NOTE: None of the univariate tests are statistically significant.

ioral variance. Without a corrective entreaty, the model accounted for only a modest 24% of the variance, but—consistent with our hypothesis—the amount of explained variance increased dramatically for respondents who were given the entreaty. Among these participants, fully 89% of the variance in the real vote was explained by intentions and perceptions of behavioral control. The difference between these two values is highly significant ($z = 4.68$, $p < .001$). To further explore the observed differences between entreaty conditions, we used the EQS software to conduct a group comparison. In the first analysis, all path coefficients were constrained to be equal for the two samples of participants (without or with a corrective entreaty). This analysis showed a good, but not perfect, fit between model and data, $\chi^2 = 17.89$, $df = 12$, $p = .12$, BBNFI = .93, RMSEA = .09; the two goodness-of-fit indices did not quite meet generally accepted criteria. In a series of subsequent analyses, successive path coefficients were allowed to vary freely. The only significant improvement was observed when the path from intention to behavior was left unconstrained. This model exhibited an excellent fit to the data, $\chi^2 = 10.96$, $df = 11$, $p = .44$, BBNFI = .96, RMSEA < .01, and the increment in the model's fit was statistically significant ($\chi^2 = 4.77$, $df = 1$, $p < .03$). Clearly, consistent with our hypothesis, exposure to a corrective entreaty had little effect on the prediction of intentions but greatly improved the prediction of actual behavior from these intentions.

Latent Disposition Hypothesis

We now consider the viability of the two hypothetical bias explanations. The latent disposition hypothesis leads us to expect that individuals who vote in favor of the disposition not only in a hypothetical situation but also under real payment conditions hold more favorable beliefs and attitudes than individuals who support the hypothetical proposition but vote against it when it is for real. To test this prediction, we examined the experimental conditions without corrective entreaty that included hypothetical as well as real referenda (i.e., Conditions 2 and 3) and selected only participants who voted in favor of the proposition in the hypothetical referendum ($N = 46$). Using multivariate analysis of variance, we compared the attitudes, subjective norms, perceived behavioral control, and intentions of participants who voted yes ($N = 27$) and participants who voted no ($N = 19$) in the real referendum. Neither the multivariate test ($F = .369$; $df = 1, 38$; $p = .92$) nor the univariate tests revealed significant differences between these groups. Table 3 displays the results of the univariate tests. It can be seen that the means of the two groups were very similar, without any consistent pattern to the differences between participants who voted no and participants who voted yes in the real referendum. There was thus no support for the hypothesis that hypothetical bias is due to a latent disposition of intermediate favorability, which would lead participants to vote yes under hypothetical conditions but no under real payment conditions.

Belief Disparity Hypothesis

Hypothetical bias. The alternative belief disparity hypothesis explains hypothetical bias by proposing that participants hold more positive beliefs about voting in favor of the referendum under hypothetical than under real payment conditions. To test this hypothesis, we compared the questionnaire responses of participants who expected to vote in a real referendum (Condition 1) to the responses of participants who expected to vote in a hypothetical referendum (Condition 2). A multivariate analysis of variance provided support for the prediction.

TABLE 4: Means and Standard Deviations for Theory of Planned Behavior Variables in Hypothetical and Real Referenda

	Hypothetical (N = *31*)		*Real* (N = *34*)		
	M	SD	M	SD	F(df = *1, 63*)
Attitude					
Direct	5.02	1.23	4.60	1.11	2.02
Belief-based	9.30	11.88	6.48	15.11	0.69
Subjective norm					
Direct	5.20	1.32	4.49	1.43	4.22*
Belief-based	16.45	21.25	1.81	18.94	8.43**
Perceived behavioral control					
Direct	5.76	1.54	5.32	1.58	1.31
Belief-based	96.67	43.73	88.71	43.62	0.53
Intention	4.88	2.04	3.81	2.17	4.25*

*$p < .05$. **$p < .01$.

In this analysis, Condition 1 versus Condition 2 was the independent factor and the theory of planned behavior measures served as dependent variables. In the multivariate analysis, responses prior to the hypothetical referendum tended, overall, to be more favorable than responses prior to the real referendum ($F = 2.63$; $df = 7, 56$; $p < .10$). More important, the univariate effects displayed in Table 4 show that the difference between hypothetical and real payment conditions was significant for intentions and that this effect could be traced primarily to differences in subjective norms. Both the direct and belief-based measures of subjective norm revealed significant differences in the expected direction between hypothetical and real referenda.

To further explore these differences, we examined the specific normative beliefs and motivations to comply that make up the belief-based measure of subjective norm. A multivariate analysis of variance revealed no significant differences for motivation to comply but significant effects for three of the four normative beliefs, as can be seen in Table 5. In comparison to hypothetical payment conditions, under real payment conditions, participants were more likely to believe that their parents and families, their friends and fellow students, and other members of their group would disapprove of their contributing $8 to the fellowship fund. There was no difference with respect to members of the university faculty and administration who were viewed as supporting a contribution in both situations.

Corrective entreaty. The belief disparity hypothesis also leads to the prediction that the effect of a corrective entreaty is to diminish or erase the differences in beliefs, attitudes, and intentions between hypothetical and real payment conditions. In a multivariate analysis of variance, the corrective entreaty had a significant multivariate main effect ($F = 2.78$; $df = 7, 118$; $p < .01$), which showed that, as expected, the corrective entreaty

TABLE 5: Differences in Normative Beliefs Between Hypothetical and Real Payment Referenda

Normative Referent	*Hypothetical* (N = *31*)		*Real* (N = *34*)		
	M	SD	M	SD	F(df = *1, 63*)
Parents and family	5.23	1.78	4.32	2.21	3.39†
Friends and fellow students	4.65	1.84	3.52	1.81	6.23*
Other group members	4.53	1.50	3.48	1.34	8.71**
Faculty and administrators	5.38	1.56	5.48	1.50	0.07

†$p < .10$. *$p < .05$. **$p < .01$.

TABLE 6: Effects of Corrective Entreaty on Theory of Planned Behavior Variables

	Without Entreaty		*With Entreaty*		
	M	SD	M	SD	F(df = *1, 124*)
Attitude					
Direct	5.39	1.57	4.60	1.37	9.18**
Belief-based	12.30	11.26	9.46	13.75	1.63
Subjective norm					
Direct	5.31	1.49	4.45	1.52	10.19**
Belief-based	17.73	21.68	5.92	16.39	11.56**
Perceived behavioral control					
Direct	5.93	1.56	5.41	1.70	3.16†
Belief-based	105.81	45.61	87.25	41.50	5.61*
Intention	5.07	2.17	3.57	2.16	15.39**

†$p < .10$. *$p < .05$. **$p < .01$.

generally lowered beliefs, attitudes, and intentions so that they were less favorable than the corresponding responses without a corrective entreaty. The relevant means and standard deviations are presented in Table 6, which also shows the results of univariate tests of the difference between participants with and without the corrective entreaty. It can be seen that the entreaty had significant effects on virtually all measures. It depressed attitudes (assessed directly) as well as subjective norms, perceptions of behavioral control, and intentions. Clearly, participants exposed to the corrective entreaty were led to form a pattern of beliefs, attitudes, and intentions that was much less supportive of a yes vote than was the case for participants who approached the hypothetical vote without the entreaty.

A second relevant piece of evidence comes from a comparison of the theory of planned behavior measures in Conditions 1 and 4. This comparison revealed that the beliefs, attitudes, and intentions of participants exposed to the corrective entreaty prior to the hypothetical vote (Condition 4) were virtually identical to the beliefs, attitudes, and intentions of participants in the real referen-

dum (Condition 1). The multivariate effect of condition on attitude, subjective norm, perceived behavioral control, and intention was not significant ($F = 0.62$; $df = 7, 55$; $p = .74$), and neither were any of the univariate effects. Taken together, these findings support the hypothesis that the corrective entreaty tends to reduce the favorability of responses under hypothetical payment conditions and to bring them in line with responses in the real referendum.

DISCUSSION

Observed discrepancies between stated intentions and actual behavior have been of concern to investigators for a long time, and they also pose a problem for the validity of contingent valuation estimates of willingness to pay. In the present study, we examined possible reasons for hypothetical bias and for the ability of an intensive corrective entreaty to ameliorate the tendency to overestimate willingness to pay. Previous studies have demonstrated the hypothetical bias, have explored its boundary conditions, and have tried to correct for it. To the best of our knowledge, this is the first study that tried to test alternative explanations of the hypothetical bias. The results suggest that the bias cannot be attributed to an insufficiently polarized latent disposition, the explanation offered by Campbell (1963) to account for inconsistencies between intentions and actions. This latent disposition hypothesis would lead us to expect that participants who vote yes in hypothetical as well as real referenda hold more favorable beliefs, attitudes, and intentions with respect to the proposition than participants who display an inconsistent pattern of behavior. However, a comparison of individuals who displayed the hypothetical bias with individuals who did not revealed little difference in their beliefs, attitudes, and intentions.

It is of course possible that in a different context, enacting a behavior has a higher threshold than expressing an intention, in support of the latent disposition hypothesis. Another way to "salvage" this hypothesis is to assume that hypothetical bias is related to facets of attitude strength other than polarity of valence (see Krosnick, Boninger, Chuang, Berent & Carnot, 1993). Thus, it is feasible that participants who do and do not vote in accordance with their hypothetical intentions hold equally positive attitudes toward a favorable vote but that the former hold these attitudes with greater confidence.

However, in the context of the present investigation, the data supported an alternative explanation offered by the belief disparity hypothesis (Ajzen & Sexton, 1999). Consistent with this hypothesis, individuals confronted with the real referendum expressed less favorable reactions with respect to voting yes than did individuals confronted with the hypothetical referendum. This finding indicates that, at least with respect to the behavior examined in this experiment, hypothetical and real contexts are qualitatively different, that people construe them in very different ways, and that questions posed in such contexts can elicit very different responses. The hypothetical situation was less likely to activate strong normative beliefs than the real payment situation. Confronted with the real payment referendum, participants seemed to realize that parents and friends might consider it unwise to give away their money, even to a worthy cause, and to believe that other group members may be opposed to making a donation and may thus not want the participant to support the proposition. It should be clear, however, that these conclusions may not be generalizable to referenda dealing with issues other than contribution to a scholarship fund or to other populations. Very different considerations may be activated for other issues and among other types of respondents. Even so, anecdotal evidence in Linn's (1965) study on racial integration also implicated normative considerations as a major distinguishing feature of hypothetical versus real behavioral situations.

The corrective entreaty developed by Cummings and Taylor (1999) was found to be very effective in reducing discrepancies between hypothetical and real votes, producing contingent value estimates that were very close to the real values. The effects of the entreaty on hypothetical votes were traced to its impact on beliefs, attitudes, and intentions with respect to voting yes in the referendum. These factors became significantly less favorable as a result of exposure to the entreaty. Of interest, the corrective entreaty was focused primarily on the consequences of a favorable vote, drawing the participants' attention to the fact that doing so would leave them with less money to spend on other things. Such an intervention would be expected to have a direct effect on behavioral beliefs and, hence, on attitudes toward the behavior. However, the data of the present study suggest that the hypothetical bias was due to a different factor: the strength of normative considerations. The corrective entreaty might therefore have missed its target, except for the fact that its impact was quite diffuse, spreading beyond attitudes to subjective norms and perceptions of behavioral control.

It may be argued, however, that the lengthy and explicit corrective entreaty produced strong demand characteristics suggesting to participants that they should lower their willingness to pay estimates and consider voting no in the hypothetical referendum. Thus, participants may simply have acceded to the experimenter's perceived expectations, without any real change in willingness to pay. Several findings argue against this possibility. For one, the corrective entreaty lowered not only willingness-to-pay estimates but also

beliefs and attitudes regarding a yes vote. There was nothing in the entreaty to suggest that people should change their beliefs and attitudes. More important, the entreaty's effects were not uniform across different normative beliefs: It lowered the perceived expectations of family, friends, and other group members but not the perceived expectations of faculty and administrators. This finding is contrary to a demand-characteristics explanation, which would predict that members of the faculty (represented by the experimenter who administered the entreaty) would have the strongest impact. Also, the entreaty did not have a uniform impact on all participants. Its impact on beliefs and attitudes was differentiated in such a manner that it greatly improved prediction of voting in the real referendum. This particular pattern of influence is inconsistent with a demand-characteristics explanation; instead, it supports the idea that the corrective entreaty led participants to form beliefs, attitudes, and intentions similar to those in a real payment situation, thereby permitting more accurate prediction of actual behavior.

The effectiveness of the corrective entreaty has important implications for influence of intentions on behavior. The assumption in the theory of planned behavior that intentions are an immediate antecedent of behavior is shared by other social psychological models (e.g., Fisher & Fisher, 1992; Gollwitzer, 1993; Triandis, 1977). Indeed, intentions are generally found to predict corresponding behavior quite well. Meta-analyses covering diverse behavioral domains have reported mean intention-behavior correlations of .47 (Armitage & Conner, 2001; Notani, 1998), .53 (Shepherd, Hartwick, & Warshaw, 1988), .45 (Randall & Wolff, 1994), and .62 (van den Putte, 1993), and in a meta-analysis of these and other meta-analyses, Sheeran (2002) reported an overall correlation of .53 between intention and behavior, suggesting that intentions account, on average, for approximately 28% of the variance in behavior. However, these meta-analyses also have shown that the magnitude of intention-behavior correlations varies widely. The present study suggests that the strength of the intention-behavior correlation is moderated by the extent to which people have realistic expectations regarding the behavioral situation. Without a corrective entreaty, intentions formed in a hypothetical context were poor predictors of behavior in a real context. In contrast, exposure to a corrective entreaty prior to making a (hypothetical) decision was shown to produce beliefs, attitudes, and intentions comparable to those that exist in a real behavioral context, and these more realistic dispositions were found to be excellent predictors of actual behavior.

These findings also have important implications for use of the theory of planned behavior in empirical research. Because this and other theories posit that intentions are an immediate precursor of behavior, and because intentions usually predict behavior quite well, many investigators have used behavioral intentions as a proxy for actual behavior. The results of the present study suggest that this practice is justified only if it can be assumed that intentions are assessed in a context that is likely to produce realistic beliefs concerning the behavior. If the beliefs that are activated when the behavior is performed differ from the beliefs that are readily accessible during questionnaire administration, actual behavior may well differ from expressed intention. Investigations that rely on intention as a proxy for actual behavior must thus be interpreted with caution.

CONCLUSIONS

The results of the present study confirm the existence of a strong bias for people to overestimate the likelihood that they will engage in a socially desirable behavior. This bias produces unrealistically high estimates of intentions to pay for a worthy cause as well as inconsistencies between intentions and actions in many other domains. One way to overcome this problem is to employ a corrective entreaty. To be effective, the entreaty must change beliefs, creating attitudes, subjective norms, perceptions of control, and intentions that are comparable to those under real payment conditions, thus reducing or eliminating the discrepancy between expressed intentions and actual behavior. As in many other applications (see Ajzen, 2001b; Armitage & Conner, 2001), the theory of planned behavior was found to be a useful conceptual framework for trying to predict and understand voting in hypothetical and real referenda. The theory permitted accurate prediction of voting intentions irrespective of the presence or absence of a corrective entreaty, but it revealed intention-behavior inconsistency in the real referendum. Whereas the corrective entreaty had no effect on the theory's ability to predict intentions, it did moderate predictive validity for real payment decisions. As expected, voting in the real referendum was predicted with considerable accuracy only for participants who had been exposed to the corrective entreaty. By creating beliefs and attitudes similar to those in the real payment situation, the corrective entreaty allowed the theory of planned behavior to make accurate predictions of voting in the real referendum.

NOTES

1. Willingness to pay for a public good also has been likened to an attitude toward the good in question (Kahneman, Ritov, Jacowitz, & Grant, 1993). Our approach focuses not on the good itself but more specifically on the intention to vote for or against a proposition to pay money for the good under consideration.

2. Participants who vote no in the hypothetical as well as in the real referendum are expected to have less favorable beliefs, attitudes, and intentions than participants who vote yes in the hypothetical referen-

dum (and yes or no in the real referendum). This difference would be predicted by both the latent disposition and the belief disparity hypotheses and can thus not be used to test the competing explanations.

3. The complete text of the corrective entreaty is available on the Internet at http://www.fs.fed.us./rm/value/cheaptalkscript.html.

4. Unlike items used to assess attitude toward the behavior directly, items assessing salient beliefs and outcome evaluations are not required to exhibit high internal consistency. Accessible behavior beliefs may contain a degree of ambivalence, with some outcomes valued positively, others negatively. There is therefore no expectation that the different beliefs will necessarily correlate highly with each other. The same logic applies to the measures of normative and control beliefs described below.

5. The questionnaire contained a few additional measures not directly related to the theory of planned behavior: a mood adjective checklist, the participants' predictions of how the other group members would vote, and their sense of moral obligation to support the petition.

6. We would like to thank Amiram Vinokur for his assistance with these analyses.

7. The analyses were repeated for all participants who were not exposed to the entreaty (Conditions 2 and 4) and those who were exposed (Conditions 3 and 5), irrespective of questionnaire order. The results were virtually identical to those reported here.

REFERENCES

Ajzen, I. (1988). *Attitudes, personality, and behavior.* Chicago: Dorsey.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior & Human Decision Processes, 50,* 179-211.

Ajzen, I. (2001a). *Constructing a theory of planned behavior questionnaire: Conceptual and methodological considerations.* Retrieved June 24, 2003, from http://www.people.umass.edu/aizen/pdf/tpb.measurement.pdf.

Ajzen, I. (2001b). Nature and operation of attitudes. *Annual Review of Psychology, 52,* 27-58.

Ajzen, I., & Driver, B. L. (1992). Contingent value measurement: On the nature and meaning of willingness to pay. *Journal of Consumer Psychology, 1,* 297-316.

Ajzen, I., & Sexton, J. (1999). Depth of processing, belief congruence, and attitude-behavior correspondence. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 117-138). New York: Guilford.

Armitage, C. J., & Conner, M. (2001). Efficacy of the theory of planned behavior: A meta-analytic review. *British Journal of Social Psychology, 40,* 471-499.

Bentler, P. M. (1995). *EQS structural equations program manual.* Encino, CA: Multivariate Software.

Blackburn, M., Harrison, G. W., & Rutstrom, E. E. (1994). Statistical bias functions and informative hypothetical surveys. *American Journal of Agricultural Economics, 76,* 1084-1088.

Blumenschein, K., Johannesson, M., Blomquist, G. C., Liljas, B., & O'Connor, R. M. (1998). Experimental results on expressed certainty and hypothetical bias in contingent valuation. *Southern Economic Journal, 65,* 169-177.

Blumer, H. (1955). Attitudes and the social act. *Social Problems, 3,* 59-65.

Brown, T. C., Ajzen, I., & Hrubes, D. (2003). Further tests of entreaties to avoid hypothetical bias in referendum contingent valuation. *Journal of Environmental Economics and Management, 46,* 353-361.

Brown, T. C., Champ, P. A., Bishop, R. C., & McCollum, D. W. (1996). Which response format reveals the truth about donations to a public good? *Land Economics, 72,* 152-166.

Campbell, D. T. (1963). Social attitudes and other acquired behavioral dispositions. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 6, pp. 94-172). New York: McGraw-Hill.

Champ, P. A., Bishop, R. C., Brown, T. C., & McCollum, D. W. (1997). Using donation mechanisms to value nonuse benefits from public goods. *Journal of Environmental Economics and Management, 33,* 151-162.

Cummings, R. G., Brookshire, D. S., & Schulze, W. D. (Eds.). (1986). *Valuing environmental goods: An assessment of the contingent valuation method.* Totowa, NJ: Rowman and Allanheld.

Cummings, R. G., Harrison, G. W., & Rutstrom, E. E. (1995). Homegrown values and hypothetical surveys: Is the dichotomous choice approach incentive-compatible? *American Economic Review, 85,* 260-266.

Cummings, R. G., & Taylor, L. O. (1999). Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method. *American Economic Review, 89,* 649-665.

Feather, N. T. (Ed.). (1982). *Expectations and actions: Expectancy-value models in psychology.* Hillsdale, NJ: Lawrence Erlbaum.

Fishbein, M. (1963). An investigation of the relationships between beliefs about an object and the attitude toward that object. *Human Relations, 16,* 233-240.

Fishbein, M. (1967). A consideration of beliefs and their role in attitude measurement. In M. Fishbein (Ed.), *Readings in attitude theory and measurement* (pp. 257-266). New York: John Wiley.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research.* Reading, MA: Addison-Wesley.

Fisher, J. D., & Fisher, W. A. (1992). Changing AIDS-risk behavior. *Psychological Bulletin, 111,* 455-474.

Fox, J. A., Shogren, J. F., Hayes, D. J., & Kliebenstein, J. B. (1998). Cvm-x: Calibrating contingent values with experimental auction markets. *American Journal of Agricultural Economics, 80,* 455-465.

Gollwitzer, P. M. (1993). Goal achievement: The role of intentions. *European Review of Social Psychology, 4,* 141-185.

Hoehn, J. P., & Randall, A. (1987). A satisfactory benefit cost indicator from contingent valuation. *Journal of Environmental Economics and Management, 14,* 226-247.

Hoehn, J. P., & Swanson, C. S. (1988). Toward a satisfactory model of contingent valuation behavior in a policy valuation context. In G. L. Peterson, B. L. Driver, & R. Gregory (Eds.), *Amenity resource valuation: Integrating economic with other disciplines* (pp. 149-158). State College, PA: Venture.

Kahneman, D., Ritov, I., Jacowitz, K. E., & Grant, P. (1993). Stated willingness to pay for public goods: A psychological perspective. *Psychological Science, 4,* 310-315.

Krosnick, J. A., Boninger, D. S., Chuang, Y. C., Berent, M. K., & Carnot, C. G. (1993). Attitude strength: One construct or many related constructs? *Journal of Personality and Social Psychology, 65,* 1132-1151.

LaPiere, R. T. (1934). Attitudes vs. actions. *Social Forces, 13,* 230-237.

Li, C. -Z., & Mattsson, L. (1995). Discrete choice under preference uncertainty: An improved structural model for contingent valuation. *Journal of Environmental Economics and Management, 28,* 256-269.

Linn, L. S. (1965). Verbal attitudes and overt behavior: A study of racial discrimination. *Social Forces, 43,* 353-364.

Loomis, J., Brown, T. C., Lucero, B., & Peterson, G. L. (1996). Improving validity experiments of contingent valuation methods: Results of efforts to reduce the disparity of hypothetical and actual WTP. *Land Economics, 72,* 450-461.

Mitchell, R. C., & Carson, R. T. (1989). *Using surveys to value public goods: The contingent valuation method.* Washington, DC: Resources for the Future.

Neill, H. R. (1995). The context for substitutes in cvm studies: Some empirical observations. *Journal of Environmental Economics & Management, 29,* 393-397.

Neill, H., Cummings, R. G., Ganderton, P., Harrison, G., & McGuckin, T. (1994). Hypothetical surveys and real economic commitments. *Land Economics, 70,* 145-154.

Notani, A. S. (1998). Moderators of perceived behavioral control's predictiveness in the theory of planned behavior: A meta-analysis. *Journal of Consumer Psychology, 7,* 247-271.

Posavac, S. S. (2001). Overbidding in value elicitation: When consumers report inflated reservation prices, and what to do about it. *Journal of Consumer Psychology, 11,* 87-97.

Randall, D. M., & Wolff, J. A. (1994). The time interval in the intention-behaviour relationship: Meta-analysis. *British Journal of Social Psychology, 33,* 405-418.

Sheeran, P. (2002). Intention-behavior relations: A conceptual and empirical review. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 12, pp. 1-36). Chichester, UK: Wiley.

Shepherd, B. H., Hartwick, J., & Warshaw, P. R. (1988). The theory of reasoned action: A meta-analysis of past research with recommendations for modifications and future research. *Journal of Consumer Research, 15*, 325-342.

Triandis, H. C. (1977). *Interpersonal behavior.* Monterey, CA: Brooks/Cole.

van den Putte, B. (1993). *On the theory of reasoned action.* Unpublished dissertation, University of Amsterdam, the Netherlands.

Received June 20, 2003
Revision accepted November 12, 2003